



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07K 14/00	A2	(11) International Publication Number: WO 98/24810 (43) International Publication Date: 11 June 1998 (11.06.98)
(21) International Application Number: PCT/EP97/06956 (22) International Filing Date: 3 December 1997 (03.12.97) (30) Priority Data: 9625283.8 4 December 1996 (04.12.96) GB (71) Applicant (for all designated States except US): JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B-9032 Wondelgem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Travessera de les Corts, 171/702a, E-08028 Barcelona (ES). DERAEMYAEKER, Marc [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VERHASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F-34000 Montpellier (FR). MAERTENS, Luc, Jacques, Simon [BE/BE]; Vier Uitersten 26, B-8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30,		B-2340 Beerse (BE). VANDEKERCKHOVE, Joel, Stefaan [BE/BE]; Rode Boukendreef 27, B-8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE). (74) Agent: BALDOCK, Sharon, Claire; Boulton Wade Tennant, 27 Fumival Street, London EC4A 1PQ (GB). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS (57) Abstract <p>Vertebrate protein homologues of UNC-53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

VERTEBRATE HOMOLOGUES OF UNC-53
PROTEIN OF C. ELEGANS

The present invention relates to vertebrate
5 homologues of UNC-53 protein of C. elegans and cDNA
sequences coding for said homologues or functional
equivalents thereof. The invention also relates to
processes for identifying compounds which control cell
behaviour, compounds identified and pharmaceutical
10 compositions containing them in addition to processes
and assays for identifying disease states in which
said gene or protein is dysfunctional.

The control of cell motility, cell shape and
directionality of cell outgrowth of axones or other
15 cell outgrowths is an essential feature in the
morphogenesis and function of both unicellular and
multicellular organisms. The control of these
processes is disturbed in a variety of disease states
in which, for example, the Receptor Tyrosine kinase
20 (RTK) signal transduction pathways, or the like, or
their downstream intra-cellular pathways (which are
shared with other extra-cellular receptors, including
cell adhesion molecules like N-CAMS and integrins) are
overstimulated.

25 Some cell surface proteins and extra-cellular
molecules controlling the directionality and potential
of cell migration have been identified, although the
processes involved are not generally understood.

It is generally considered that a long-range
30 migration of a cell process (also known as a growth
cone extension) is a stepwise event, whereby prior to
and after each extension there is the formation of a
structure at the leading edge of the cell which senses
signals in the environment instructing the cell to
35 either stabilise a cell process extending in a

- 2 -

preferred direction, or to cause a lamellipodium to extend a process in a given direction. Localised stabilisation of the actin cytoskeleton and association with plus end regions of microtubules is a general cell biological process underlying the choice of directional extension. Microtubule binding directing these processes has not previously been identified. The present inventors have surprisingly found that UNC-53 protein of *C. elegans* and vertebrate homologues thereof is involved in binding of microtubules and particularly of plus end regions of microtubules.

A gene from the free-living nematode Caenorhabditis elegans designated "unc-53" has been previously identified and cloned (Abstract, International C. elegans Meeting, June 1-5 1991, Madison, Wisconsin, 58, Bogaert and Goh). The present inventors previously identified UNC-53 protein as a signal transducer or signal integrator controlling the directionality of cell migration and/or cell shape in C. elegans (WO 96/38555). Increased UNC-53 protein activity was found to be proportional to cell process extensions in the correct direction of cell migration. The unc-53 gene was found to encode a signal transduction molecule that transduces a signal from an RTK such as, for example, via the adaptor protein SEM-5/GRB-2, to the machinery controlling directional growth cone extension or stabilisation, in a highly dosage - dependent fashion.

Genetic and experimental analysis of C. elegans UNC-53 mutants showed that mutations in the unc-53 gene do not affect the general ability of cells to migrate but rather affect the ability of cells to migrate under specific antero-posterior cues. Reduction of UNC-53 activity leads to loss of

- 3 -

direction and reduction of growth cone extension as indicated by the directionality of random extension cycles observed in excretory canal growth cones in UNC-53 mutants.

5 The function of UNC-53 is highly sensitive to its dosage or activity. Reduction of function leads to proportional reduction of migration to the specific signal while increased expression, using transgenic expression of UNC-53 in muscle cells, leads to
10 increased directional migration. The data lead to the conclusion that UNC-53 functions as an integrator of a directional signal in the organism whereby reception of signals leads to growth cone extension in the correct direction.

15 Certain alleles of UNC-53 enhance the sex myoblast migration defect of SEM-5 C. elegans mutants in a receptor tyrosine kinase signal transduction pathway (Stern et al 1993 mol. Biol. cell, 4, 1175-1188). While the genetics suggests that UNC-53 and
20 SEM-5 cooperate to regulate sex myoblast migration, genetic experiments do not permit a conclusion that this is the result of a direct molecular interaction. The inventors previously identified a potential sem-5/GRB-2 binding site and showed in two types of
25 biochemical experiments that UNC-53 physically interacts with SEM-5. The present inventors conclude that UNC-53 encodes a signal transduction molecule that transduces extracellular signals for directional migration via the adapter protein SEM-5/GRB-2 to the
30 machinery controlling directional growth cone extension or stabilization.

 Several lines of evidence indicate that UNC-53 might act as an adapter linking extracellular signals to the actin cytoskeleton. Firstly, UNC-53 has shown
35 homology to cortical actin binding proteins and that

- 4 -

it is capable of binding F-actin in vitro. In addition expression of UNC-53 in mammalian cells leads to changes in the F-actin cytoskeleton. Very low levels of UNC-53 expression increase the number of
5 filopodia and actin microspikes protruding from the cell surface. Cells expressing UNC-53 also exhibit increased neurite extension and increased cell motility. UNC-53 thus also acts as an activator of migration.

10 Considering all available data the following possible mechanisms of action of UNC-53 can be formulated.

The choice and activation of directional growth cone extension can be accounted for by local
15 activation of UNC-53 via a SEM-5/GRB2 complex to a receptor (eg receptor tyrosine kinase signal) which reads a localized or directional signal. Changes in growth cone steering are preceded by the formation of a localized actin patch in the area of the growth cone
20 receiving the highest signal (Bentley and O'Connor et al. Curr. Op. NeuroBiol. 1994, vol 4, 43-48). UNC-53 might be directly involved in forming these actin patches through its own actin binding or cross-linking properties. Alternatively activated UNC-53
25 may (eg via its nucleotide binding domain) transduce a signal to as yet unidentified effectors. For example, activation of the small GTP-binding protein cdc42 or a related protein leads to formation of small actin patches as well as the formation of small filopodia.
30 The unc-53 pathway may be upstream of cdc42 or both signal transducers might share downstream pathways.

The present inventors thus decided to investigate if a similar protein was present in higher organisms such as vertebrates.

35 The present inventors describe the identification

- 5 -

of a family of genes in vertebrates, and particularly in man and mouse with extensive structural homology to UNC-53. The present inventors have surprisingly found that the nucleotide domains of UNC-53 from C. elegans and UNC-53 from vertebrates similarly activate motility, establishing functional equivalence. Furthermore these domains are shown to be capable of transforming NIH3T3 cells in vitro. The inventors also found changes in RNA transcripts in transformed cell lines compared to normal human tissues suggesting a role for UNC-53 in cell differentiation, morphogenesis and disease. Furthermore, in vitro assays and transgenic models are also described that identify pharmacological modulators of UNC-53 activity and assays to identify proteins interacting with UNC-53.

According to a first aspect of the present invention, there is provided a vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein homologue comprises an amino acid sequence having a statistically significant homology to the UNC-53 protein of C. elegans as illustrated in figure 2. According to the present invention a derivative should be taken to mean mutational derivatives, fusions, internal deletions, splice variants and muteins.

There is also provided according to a second aspect of the present invention a vertebrate protein homologue of UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence homology blocks A, B, C, D or E as illustrated in Figure 9a, or block F in Figure 12a or a sequence having a statistically significant homology therewith.

- 6 -

Preferably, said vertebrate homologue is a human protein or a mouse protein.

According to a further aspect of the invention there is provided a vertebrate protein homologue of an
5 UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E or F which differ from those blocks of Figure 9a and Figure 12a to a significant extent only in conservative amino acid changes. In an
10 even further aspect of the invention there is provided a vertebrate protein having an amino acid sequence encoded by the nucleotide sequence from position 1 to position 6013 as illustrated in Figure 9b. There is also provided a vertebrate protein having an amino
15 acid sequence encoded by the nucleotide sequence illustrated in Figure 11d, or a functional equivalent derivative, or bioprecursor of said homologue.

According to a further aspect of the present invention there is provided a vertebrate protein
20 having an amino acid sequence corresponding to the prosite signatures as illustrated in Figure 28 for each of said homology blocks as defined above. Advantageously the prosite signatures can be used to identify a protein having a statistically significant
25 homology to the UNC-53 protein of C. elegans. (Luethy et al 1994, Protein Science, 3, 139-146).

A further aspect of the invention comprises a vertebrate homologue according to the invention comprising an amino acid sequence as shown in figure
30 9b or 11d or an amino acid sequence which differs from the amino acid sequences shown in these figures to a significant extent only in one or more conservative amino acid changes.

In a further aspect of the present invention
35 there is also provided a nucleic acid molecule, which

- 7 -

is preferably DNA, and which encodes a vertebrate
homologue of UNC-53 protein of C. elegans, or a
functional equivalent derivative, fragment or
bioprecursor of said homologue according to the
5 invention. Preferably, the cDNA comprises a sequence
of nucleotides encoding an amino acid sequence as
illustrated in figures 9b or 11d or an amino acid
which differs from the sequences shown in these
figures to a significant only in one or more
10 conservative amino acid changes. Preferably the DNA is
cDNA, which cDNA comprises at least from position 1
to 6013 of the sequence shown in Figure 9b.
Alternatively the cDNA may comprise the sequence
illustrated in Figure 11d. Also provided by the
15 present invention is a nucleic acid sequence capable
of hybridising to the nucleic acid or DNA sequences
according to the invention under high stringency
conditions, which conditions are well known to those
skilled in the art.

20 The cDNA according to the invention may be
included in an expression vector which may itself be
used to transform or transfect a host cell, which cell
may be bacterial or eukaryotic in origin including
such as, for example an animal or plant cell a fungal
25 cell or an insect cell. Thus, advantageously, once
the cDNA corresponding to the genome of the vertebrate
homologue of UNC-53 of C. elegans is synthesised,
using for example, reverse transcriptase or the like,
a range of cells, tissues or organisms may be
30 transfected following incorporation of the selected
cDNA clone into an appropriate expression vector. The
expression vector according to the invention may
comprise a promoter of C. elegans or one of human
mouse or viral origin and optionally a sequence
35 encoding a reporter molecule, such as, for example,

- 8 -

green fluorescent protein.

The present invention, therefore, also further comprises a transgenic cell, tissue or organism comprising a transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment derivative or bioprecursor of said homologue. The term "transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans" as used herein means a suitable nucleic acid sequence which leads to the expression of a vertebrate homologue of UNC-53 protein of C. elegans having the same function and/or activity. The transgene may include, for example, genomic nucleic acid isolated from the appropriate vertebrate or synthetic nucleic acid including cDNA. The term "transgenic organism, tissue or cell, as used herein means any suitable organism and/or part of an organism, tissue or cell, that contains exogenous nucleic acid either stably integrated in the genome or in an extrachromosomal state.

Preferably the transgenic cell comprises any of, a COS cell, HepG2 cell, MCF-7 or N4 neuroblastoma cell or a NIH3T3 cell or a colorectal or carcinoma cell or a human derived cell such as a fibroblast or the like. The transgenic organism may be an insect, a non-human animal or a plant and preferably C.elegans or a related nematode. Preferably, the transgene comprises the nucleic acid sequence encoding the vertebrate homologue or a functional fragment of said gene according to the invention as described above. The transgene preferably comprises an expression vector according to the invention.

The term "functional fragment" as used herein should be taken to mean a fragment of the gene coding for the vertebrate homologue of the UNC-53 protein of

C. elegans or a functional equivalent or derivative or bioprecursor of said protein. For example, the gene may comprise deletions or mutations but may still encode a functional vertebrate homologue of UNC-53 protein.

Further provided by the present invention is a method of producing a mutant vertebrate non-human organism or cell having a mutation in the wild-type gene coding for the vertebrate homologue of UNC-53 protein, which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes located thereon, which method comprises inducing a mutation in the vertebrate homologue of UNC-53 protein in said organism or cell. These mutant organisms or cells may be used in a screen to identify the effects of compounds on these cell functions.

The vertebrate homologue of UNC-53 protein of C. elegans or the cDNA or genomic DNA encoding it or a functional equivalent, derivative, fragment or bioprecursor of said homologue, may advantageously be used as a medicament, or in the preparation of a medicament to promote neuronal regeneration, revascularisation or wound healing or the treatment of chronic neurodegenerative disorders or acute traumatic injuries or fibrotic disease or physiological events requiring the polarity of cells or epithelia. The present inventors have also found that the vertebrate homologue of UNC-53 protein plays a role in a transformed state of cells. Accordingly, the vertebrate homologue, dominant positive or negative mutants thereof, or inhibitors thereof may advantageously be used to induce or alleviate contact inhibition in a cell or in preventing cancer

- 10 -

development. Typically, the above medical conditions may be treated in mammals and more preferably humans by either a homologue of UNC-53 protein or alternatively by a nucleic acid coding for such a protein. Alternatively an antisense oligonucleotide to said UNC-53 homologue may be used to prevent its expression. Examples of other nucleic acid sequences which may be used include 3' untranslated regions of mRNA which could be used to prevent transcription of the genomic sequence encoding for the vertebrate homologue of UNC-53 protein.

The vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said protein may be incorporated into a pharmaceutically acceptable composition together with a suitable carrier, diluent or excipient therefor. The pharmaceutical composition may advantageously comprise, additionally or alternatively, the nucleic acid sequence according to the invention as defined above.

The present invention also provides for a method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, transformation, cell shape or motility or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes thereon which method comprises contacting said compound with a transgenic cell according to the invention and screening for a phenotypic change in said cell. Preferably the method can determine whether the compound comprises an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which pathway said vertebrate homologue of UNC-53 protein, or a functional equivalent, derivative, fragment or bioprecursor of

- 11 -

said vertebrate homologue is a component or whether said compound is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell. The present invention also provides a method to determine that the protein in said signal transduction pathway is a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

Preferably, the phenotypic change to be screened comprises a change in cell shape or a change in cell motility. Where a transgenic cell is used in accordance with one embodiment of the method of the invention, an N4 neuroblastoma cell may be used and in such an embodiment the phenotypic change to be screened may be the length of neurite growth or changes in filipodia outgrowth or alternatively changes in ruffling behaviour or cell adhesion or any change in microtubule cytoskeleton or any change in localisation of proteins on plus end regions of microtubules or any change in cell death such as in apoptosis. In an alternative embodiment of the method of the invention, the transgenic cell may comprise an MCF-7 breast cancer cell. Typically in such an embodiment the phenotypic change to be screened comprises the extent of phagokinesis or filipodia formation. In an alternative embodiment of this aspect of the invention, the transgenic cell may comprise an NIH3T3 cell. Typically in such an embodiment the phenotypic change to be screened comprises loss of contact inhibition of foci formation. The method according to the invention, may also utilise a mutant cell or mutant organism according to the invention as described above, where the mutant cell is capable of growing in tissue

- 12 -

culture or in vivo and either of which cell or organism has a mutation in the wild-type unc-53 gene.

5 In accordance with the present invention, a "phenotypic change", may be any phenotype resulting from changes at any suitable point in the life cycle of the cell, tissue or organism defined above, which change can be attributed to the expression of the transgene such as for example, growth, viability, morphology, behaviour, movement, cell migration or
10 cell process or growth cone extension of cells and includes changes in body shape, locomotion, chemotaxis, contact inhibition, mating behaviour or the like. The phenotypic change may preferably be monitored directly by visual inspection of the cell as
15 a whole or particularly by monitoring the F-actin cytoskeleton microtubule network and plus end stability of microtubules or proteins thereon or alternatively by for example measuring indicators of viability including endogenous or transgenically
20 introduced histochemical markers or other reporter genes, such as for example β -galactosidase or green fluorescent protein.

A compound which is identifiable by the method according to the invention as described above, as an
25 enhancer of the processes identified above such as the regulation of cell shape or motility or the direction of cell migration may be used as a medicament, or alternatively in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or
30 wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease. Examples of promoting neuronal regeneration include, for example, peripheral nerve regeneration after trauma and spinal cord trauma.

35 Where a compound is identified in accordance with

- 13 -

the method described above as being an inhibitor of the regulation of cell shape etc., the compound may be used as a medicament, or in the preparation of a medicament, for substantially alleviating spread of disease inducing cells, such as in spread of cancer, or the like in metastasis or in alleviating loss of contact inhibition. Advantageously, any of the compounds which may have been identified as an inhibitor or an enhancer in accordance with the method as described above, may also be included in a pharmaceutical composition comprising the respective compound and a pharmaceutically acceptable carrier, diluent or excipient therefor.

The particular mechanism of action of a compound identified as either an inhibitor or an enhancer of the cell motility shape, growth or direction of cell migration or microtubule association or to the plus end region thereof is not limiting. Preferably the compound acts as an inhibitor or enhancer of a signal transduction pathway. The compound may also act on a parallel pathway or directly on the vertebrate homologue of UNC-53 protein of C. elegans. For example, the method of action of the compound may include direct interaction with the vertebrate homologue of UNC-53 protein, interaction with processes for regulating phosphorylation or dephosphorylation of the vertebrate homologue of UNC-53 or with processes regulating activity of an unc-53 gene or with processes for post-transcriptional or post-translational modification or the like.

Preferably the compound is identified by the method according to the invention as an inhibitor or an enhancer, by utilising differences of phenotype of the cell, tissue or organism, which are visible to the eye. Alternatively indicators of viability including

- 14 -

endogenous or transgenically introduced histochemical markers or a reporter gene may be used.

According to a further aspect of the invention there is also provided a transgenic cell or tissue
5 culture which has been constructed to comprise a promoter sequence of a gene coding for a vertebrate homologue of UNC-53 of C. elegans or a functional equivalent, derivative fragment, or bioprecursor of said homologue operably linked to a nucleic acid
10 sequence encoding a reporter molecule. Preferably, the reporter sequence encoding the reporter molecule which comprises a detectable protein, for example one which may be monitored by eye inspection such as antibiotic resistance, β -galactosidase or a molecule
15 detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also provides a method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a
20 vertebrate homologue of UNC-53 protein in C. elegans, or a functional equivalent, derivative fragment or bioprecursor of said homologue, which method comprises the steps of:

- 25 (a) contacting said compound with a transgenic cell according to the invention as described above,
- (b) monitoring the level of said reporter molecule and comparing results obtained from this monitoring step with a control comprising a
30 transgenic cell having the promoter sequence of a gene coding for a vertebrate homologue of UNC-53 protein, or a functional fragment of said homologue and the reporter molecule, in the absence of the compound.

35 In one embodiment of the method according to this

- 15 -

aspect of the invention the reporter molecule may comprise messenger RNA.

5 A compound identified as an enhancer of transcription of the gene coding for the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor of said homologue may also be used as a medicament, or in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound
10 healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Furthermore, such compounds may be included in a pharmaceutical composition including a pharmaceutically acceptable carrier, diluent or
15 excipient therefor. Any compounds identified as inhibitors of transcription may, advantageously, be used in alleviating the spread of disease inducing cells such as cancers or metastasis or loss of contact inhibition.

20 The present invention also provides a kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell growth, transformation, cell motility or shape or the direction of cell migration which kit comprises at
25 least one transgenic or mutant cell or transgenic or mutant non-human organism according to the invention as described above and a plurality of wild-type cells or one organism of the same type, or a cell line or tissue culture and means for contacting said compound
30 with said cell or organism.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans
35 or a functional equivalent, derivative or fragment

- 16 -

thereof, which kit comprises at least one transgenic cell or cells according to the invention and means for contacting said compounds with said cells.

For the purposes of the present invention, the
5 term "gene coding for a vertebrate homologue of UNC-53
or a functional fragment of said homologue" includes
the nucleic acid sequence shown in Figures 9b or 11d
or a fragment thereof, including the differentially
10 spliced isoforms and transcriptional starts of the
nucleic acid sequence and which sequence encodes a
vertebrate homologue of UNC-53 protein or a functional
equivalent, derivative, fragment or bioprecursor of
the protein.

The present invention also provides methods of
15 identifying genes of vertebrates or fragments of said
genes, which encode proteins which are active in the
signal transduction pathway of which the vertebrate
homologue of UNC-53 is a component. A preferred
method comprises hybridizing to an appropriate cDNA
20 library a nucleotide sequence, as defined herein, or a
fragment thereof under appropriate conditions of
stringency in order to identify genes having
statistically significant homology with the cDNA
clones of any one of the cDNA sequences according to
25 the invention described above.

Furthermore, there is also provided by the
present invention a method of identifying a protein
which is active in the signal transduction pathway of
a cell of which a vertebrate homologue of UNC-53
30 protein of C. elegans or a functional equivalent,
fragment or bioprecursor of said vertebrate homologue
is a component. According to this aspect of the
invention, the method comprises;

(a) contacting an extract of said cell with an
35 antibody to the vertebrate homologue of UNC-53

- 17 -

protein or a functional equivalent, fragment or bioprecursor of said protein,

(b) identifying the antibody/vertebrate homologue of UNC-53 complex, and

5 (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the antibody.

The vertebrate homologue of UNC-53 protein, therefore may bind regions of other proteins involved
10 in the signal transduction pathway. It is also possible to sequentially identify a whole range of proteins involved in the signal transduction pathway.

Antibodies to the vertebrate homologue of UNC-53 protein may be produced according to known techniques
15 as would be known to those skilled in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with a protein or epitope of a protein according to the invention and recovering immune serum.

20 This aspect of the invention further comprises a method of identifying a further protein or proteins which are active in the signal transduction pathway of a cell of which UNC-53 is a component which method comprises:

25 (a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein in the method as described above,
(b) contacting a cell extract with the antibody,
(c) identifying the antibody/protein complex,
30 (d) analysing the complex to identify any further protein bound to the first protein other than the antibody, and
(e) optionally repeating steps (a) to (d) to identify further proteins in the pathway.

35 According to this aspect of the present

- 18 -

invention, the antibody starts the process by binding to the vertebrate homologue of UNC-53 protein or a functional equivalent thereof in the signal transduction pathway. Any other proteins found
5 complexed to the bound antibody or UNC-53 protein can then be used to identify further interacting proteins involved in the pathway.

It may also be possible to identify proteins involved in the signal transduction pathway of a cell
10 of which the vertebrate homologue of UNC-53 or a functional equivalent derivative or bioprecursor thereof is a component by using a vertebrate homologue of UNC-53 protein of C. elegans. According to this aspect of the invention the method comprises:

- 15 (a) contacting an extract of the cell with the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue,
- (b) identifying the vertebrate homologue of
20 UNC-53 protein/protein complex formed and
- (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein

25 This method can also advantageously be used to identify further proteins in a signal transduction pathway of a cell by contacting an extract of the cell used as described above, with any protein identified from step (c) above not being a vertebrate homologue
30 of UNC-53 protein and repeating steps (b) and (c).

Other methods which may be used for identifying proteins in a signal transduction pathway of a cell may comprise for example a western blot overlay method which method is well known to those skilled in the
35 art. Cell extracts are run on gels to separate out

- 19 -

protein and subsequently blotted onto a nylon membrane. These membranes may then be incubated, for example in a medium containing a vertebrate homologue of UNC-53 having a label attached thereto such as a biotin or radiolabel and any protein conjugates visualised with for example a streptavidin or alkaline phosphatase conjugated antibody.

The present invention also advantageously provides a process for the preparation of binding antibodies which recognise proteins or fragments thereof involved in the rate and direction of cell migration or the control of cell growth or shape, for the above methods.

The monoclonal antibody for binding to the appropriate vertebrate homologue of UNC-53 (or its functional equivalent) may be prepared by known techniques as described by Kohler R. and Milstein C., (1975) Nature 256, 495 to 497.

Another method which may be used to identify proteins involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent or derivative or bioprecursor is a component involves investigating protein-protein interactions using the two-hybrid vector method. This method is well known to those skilled in the art and which was first developed in yeast by Chien et al (1991). This technique is based on functional reconstitution in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA

- 20 -

sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

15 An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example a sequence coding for the vertebrate homologue of UNC-53. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein, preferably from the signal transduction pathway of the vertebrate in question. Any interaction between the vertebrate homologue of UNC-53 protein and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter

- 21 -

molecule such as β -galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes. This method enables any interactions between proteins involved in the signal transduction pathway or a parallel or redundant pathway to be investigated.

Any proteins identified in the signal transduction pathway of the cell, which may be for example a mammalian cell, may also be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention also provides a process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, or derivative of the protein, which process comprises culturing the cells transformed or transfected with a cDNA expression vector having any of the cDNA sequences according to the invention as described above, and recovering the expressed vertebrate homologue of UNC-53 protein. The cell may advantageously be a bacterial, animal, insect or plant cell.

A particularly preferred process for producing a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or fragment of said homologue comprises using insect cells. Accordingly, the invention provides a process for producing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of the UNC-53 protein, which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a nucleotide vector encoding the vertebrate homologue of UNC-53 protein or a functional equivalent, fragment

- 22 -

or bioprecursor thereof downstream of the Baculovirus polyhedrin promoter and recovering the expressed protein. Advantageously, this method produces large amounts of protein for recovery. The insect cell may
5 be from for example Spodoptera frugiperda or Drosophila Melanogaster.

In accordance with the present invention, a defined nucleic acid sequence includes not only the identical nucleic acid but also any minor base
10 variations from the natural nucleic acid sequence including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid), due to the degenerate code in conservative amino acid substitution. The
15 term "nucleic acid sequence" also includes the complimentary sequence to any single stranded sequence given which includes the definition above regarding base variations.

Furthermore, a defined protein, polypeptide or
20 amino acid sequence according to the invention, includes not only the identical amino acid sequence but also minor amino acid variations from the natural amino acid sequence including conservative amino acid replacements (a replacement by an amino acid that is
25 related in its side chains). Also included are amino acid sequences which vary from the natural amino acid but result in a polypeptide which is immunologically identical or similar to the polypeptide encoded by the naturally occurring sequence. Such polypeptides may
30 be encoded by a corresponding nucleic acid sequence.

A further aspect of the invention provides a nucleic acid sequence of at least 15 nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides.

35 These sequences may, advantageously be used as

- 23 -

probes or primers to initiate replication or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used in diagnostic kits or the like for detecting for the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with a sample under hybridising conditions and detecting for the presence of any duplex formation between the probe and any nucleic acid in the sample. Nucleic acid sequences according to the invention may also be produced using recombinant or synthetic means such as described in Sambrook et al (Molecular Cloning: A Laboratory Manual, 1989). Advantageously, human allelic variants or polymorphisms of the DNA according to the invention may be identified by, for example, probing DNA libraries from a range of individuals for example from different populations. Furthermore, nucleic acids and probes according to the invention may be used to sequence genomic DNA from patients using techniques well known in the art, such as the Sanger Dideoxy chain termination method, which may advantageously ascertain any predisposition of a patient to certain proliferative disorders.

A method of detecting whether a compound is an inhibitor or an enhancer of expression of a vertebrate homologue of UNC-53 of C. elegans, or a functional equivalent, derivative or fragment of said vertebrate homologue is also provided which method comprises contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

Preferably the cell is a transgenic cell as described above. Alternatively the cell may have

- 24 -

undergone loss of contact inhibition.

The present method also provides for determining whether said compound is an inhibitor of expression of said vertebrate homologue. In one embodiment the compound to be tested comprises a nucleic acid.

Preferably said nucleic acid sequence comprises an antisense DNA sequence or a mRNA sequence.

Preferably said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

Alternatively, the compound to be tested may be a protein. Preferably, said protein comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue and preferably comprises a protein identified by the methods as described herein.

The present invention also provides a pharmaceutical composition comprising a compound, for example an antisense nucleic acid identified according to the above described method together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A nucleic acid sequence or protein identified according to this aspect of the invention may be used as a medicament, or in the preparation of a medicament, for treating loss of contact inhibition or cancer which is mediated by a vertebrate homologue of UNC-53 protein or a functional equivalent, fragment, derivative or bioprecursor of said homologue.

Further provided by the invention is a nucleic acid as defined above for use in preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said homologue.

- 25 -

According to a further aspect of the invention there is provided a plasmid pCB201 deposited under LMBP Accession No. LMBP 3594 and a MCF-7 and a NIH/3T3 cell line transfected with plasmid pCB201 deposited
5 under LMBP Accession Nos. LMBP 1601 CB and LMBP 1603 CB respectively. Further provided by the invention is phage lambda 3b coding for Hu-UNC-53/1 and deposited under Accession No. LMBP 1604CB (or 3595). Also provided are plasmids pLM1 deposited under Accession
10 No. LMBP 3762, pLM4 (LMBP 3763), pEGFP72 (LMBP 3764) and pCB501 (LMBP 3765). Further provided is a Bac clone comprising a fragment of hu-unc-53/2 gene (LMBP 3773) and a worm strain comprising a chimeric C.elegans human unc53 gene deposited under LMBP
15 Accession No. LMBP-1663CB.

Further provided by the invention is an assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract
20 thereof with an antibody to said vertebrate homologue, or a functional equivalent, derivative or bioprecursor thereof, which antibody is fused to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

25 Preferably the reporter molecule is an antibody conjugated to for example a flurophore such as fluorescein or alternatively to an enzyme such as strepavidin.

There is also provided a method for detecting for
30 expression of a gene coding for a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or
35 corresponding to said vertebrate homologue or a

- 26 -

functional equivalent, fragment or bioprecursor thereof with a cell extract, which probe is linked to a reporter and analysing for the presence of said reporter.

5 Preferably the probe is a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

10 Preferably the complimentary sequence is a 3' or 5' untranslated region of said mRNA. Preferably said reporter may be a dig label, a fluorophore, a hapten or a radiolabel.

15 Alternatively said probe comprises an antibody specific for said vertebrate homologue of said UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor therefor.

20 Preferably the reporter is an antibody conjugated to for example a fluorophore such as fluorscein or alternatively an enzyme such as streptavidin.

25 As described above UNC-53 protein of C.elegans has been found to localise to microtubule and particularly to microtubule (+) ends. Therefore, there is provided by a further aspect of the present invention a method of determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to any of claims to 1 to 9 to microtubules or plus end regions thereof, which method comprises (a) contacting
30 said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule (b) screening for the localisation of said reporter molecule as compared to a cell
35 according to step (a) which has not been contacted

- 27 -

with said compound.

A compound identifiable by the above method also forms part of the present invention. Such a compound identified as an inhibitor of localisation or
5 association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition. Further a compound identified as an enhancer of
10 association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in for example promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic neurodegenerative diseases or
15 acute traumatic injuries or fibrotic disease. These compounds may then be included in a pharmaceutical composition, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Also provided by the present invention is a kit
20 for determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to the invention with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing
25 UNC-53 and a reporter molecule or a host or transgenic cell according to the invention and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells. Compounds identified as
30 inhibitors or enhancers or microtubule association described above may advantageously be included in a composition and linked to unc-53 protein of C.elegans or a vertebrate homologue thereof according to the invention to target the compounds to the microtubules
35 or the plus end regions thereof. Such a composition

- 28 -

may also comprise, for example, a suitable transfecting or transformation agent.

According to a further aspect of the invention there is provided a method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to the invention, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting said protein to said microtubule or a plus end region thereof. An even further aspect of the invention comprises a method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue thereof according to the invention, which method comprises a) contacting either an extract from a cell or cells expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising candidate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein, b) identifying any covalently modified UNC-53 protein from step a) and c) identifying said molecule involved in said modification step. Such an indicator may be ³²P.

Further provided by the invention is a method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to the invention, or which alleviates or enhances apoptosis. The method of the former comprises contacting said compound with a transgenic cell, tissue or organism according to the invention and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof. In the case of apoptosis the

- 29 -

method comprises monitoring the effect of the compound on cell death.

The invention may be more clearly understood from the following examples which are only exemplary, with reference to the accompanying drawings wherein

Figure 1 illustrates the sequence of plasmid pTB72 which codes for the full length UNC-53 protein in C. elegans, deposited under LMBP Accession No. 3486.

Figure 2 illustrates the full-length UNC-53 protein from C. elegans.

Figure 3 is a Tblastn search of the EST division of Genbank with the ORF of the longest known Ce-UNC-53 cDNA. tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in Ce-UNC-53.

Figure 4 illustrates a search of the Genbank databases with part of the nucleotide binding domain of Ce-UNC-53. It does not identify statistically significant proteins except for the C. elegans cosmid containing Ce-unc-53.

Figure 5 illustrates a three frame translation of EST gb:R41071.

Regions of homology with Ce-Unc-53 in two different frames are underlined. The spacing between the blocks of homology is of similar size to that in Ce-UNC-53. Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with Ce-UNC-53 (see alignment in fig. 12).

Figure 6 is a BLASTN search of the EST division of Genbank with Hu-unc-53/1 cDNA cosmid 3b.

Figure 7 is a TBLASTN search of the Genbank sequence database with the 961 amino acid ORF of cDNA 3b of hu-UNC-53/1 : hu-UNC-53/1 forms a unique pair with Ce-UNC-53 (cosmid F45E10) compared to the rest of

- 30 -

the database.

Figure 8 is a diagram illustrating the length and overlap and tissue source of the different cDNA clones of the 3' end of Hu-UNC-53/1 isolated in this work.

5 Figure 8a. is a diagram illustrating the further sequence of the Hu-UNC-53/1 and overlap of constructs to obtain the further sequence.

Figure 8b is a diagram illustrating the 3' end of Hu-UNC53/1 and the EST clones present in the
10 database.

Figure 9a is an annotated sequence listing of clone 3b of hu-UNC-53/1 including the EcoR1 polylinker GAATTC. The predicted Open Reading Frame of Hu-UNC-53/1 is listed below the sequence. Blocks A B C D and
15 E which are similiar to Ce-UNC-53/1, a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2 and the 3' untranslated leader sequences are marked with arrows and labelled.

Figure 9b is an annotated sequence listing of
20 Hu-UNC-53/1 available at this moment. The predicted Open Reading Frames of Hu-UNC-53/1, pLM1, pLM3, pLM4, pCB251, pLM5 and pCB201, the homology blocks A,B,C,D and E, the position of a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2, the position of
25 phh14-3, pCB212, pCB210-14, phh3b, phh15, the position of the reverse primers HU53rv1, HU53rv2, HU53rv3 and HU53rv4, the position of peptides B72628 (=28/1), B72627, B72626 and B72625 are listed below the sequence.

30 Figure 10 is an annotated sequence listing of the insert of clone gbAA049124 (EST479167) of mu-UNC-53/1. The open reading frame and 3' untranslated sequence is marked with an arrow.

35 Figure 11a is an annotated sequence listing of the insert of clone gbH09036 (EST46037) of Hu-UNC-

- 31 -

53/2.

Figure 11b is a novel DNA sequence of HU-UNC-53/2 extended by RT-PCR. This DNA sequence is not present in EST-46037 and extends the ORF beyond position 1109 of Figure 11a to an ORF from position 18 to 1793.

Figure 11c summarises how the 3' and 5' extensions of hu-unc-53/2 were made.

Figure 11d compiles the sequence of hu-unc-53/2. The boxed sequences are the primer sequences used for the respective extension steps described in the experimental methods section.

Figure 11e illustrates the sequences of the extensions summarised in figure 11c.

Figure 11f illustrates the sequence information illustrating four alternative Start sites observed for hu-unc-53/2.

Figure 12 is an illustration of a Tblastn search of the EST division of Genbank with 680aa starting at the C-terminus of the alpha actinin domain of hu-unc-53/2.

Figure 12a is an illustration of an amino acid alignment of the available sequence of C.elegans unc-53 and hu-unc-53/1 and hu-unc-53/2.

12b. is an illustration of similarity plots for Ce-unc-53 and hu-unc-53/1 (top) and for hu-unc-53/1 and hu-unc-53/2.

Figure 13 is an annotated sequence listing of expression vector pCB201 containing homology block E from Hu-UNC-53/1 cloned in a pCDNA3.1-HIS expression vector. The HIS and T7-tags, PCR primer used to modify hu-UNC-53/1 and ORF are marked.

Figure 14 is a diagram showing the alignment of the homologous regions of hu-UNC-53/1 and mu-UNC-53/1.

Figure 15 is an annotated sequence listing of expression vector pCDU3 containing part of Ce-UNC-53/1

- 32 -

cloned in expression vector pCDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

5 Figure 16 is an annotated sequence listing of expression vector pCDU4 containing part of Ce-UNC-53/1 cloned in expression vector pCDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

10 Figure 17 is an annotated sequence listing of expression vector pCDU2 containing part of Ce-UNC-53/1 cloned in expression vector pCDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

15 Figure 18 illustrates MCF-7 cells transfected with pCB201 (upper) compared to mock transfected MCF-7 cells (phase contrast image). The control cells are spread out on the tissue culture plastic and exhibiting few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths per cell (arrowheads).

20 Figure 19 is a phase contrast image of MCF-7 cells, transfected with pCDNA3.1 (19a), pCDU4 (19b), pCDU3 (19c), pCDU2 (19d) and pTB72 (19e).

25 Figure 20 is an F-actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pCDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

30 Figure 21 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of MCF-7 cells transfected with pCDNA3.1 (21a), pCDU4 (21b), pCDU3(21c), pCDU2 (21d) and pTB72 (21e).

35 Figure 22 is a phase contrast image of N4 neuroblastoma cells transfected with pCDNA3.1 (22a), pCDU4 (22b), pCDU3 (22c), pCDU2 (22d) and pTB72 (22e).

- 33 -

Figure 23 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pCDNA3.1 (23a), pCDU4 (23b), pCDU3 (23c), pCDU2 (23d) and pTB72 (23e).

5 Figure 24 illustrates phase contrast images of small (top), medium (middle) and large foci (bottom) induced in a monolayer of NIH3T3 cells by transfection with pCB201.

10 Figure 25(c) illustrates human metaphase chromosomes probed with a probe 1p34 and figures 25a and 25b indicating the chromosomal location of hu-UNC-53/1 in 1q31. Essentially the same techniques were used to assign the gene hu-unc-53/2 to chromosome locus 11p15 (25d and e) as illustrated in micrograph
15 25f.

 The ideograms 25a and 25d are from the International System for Human Cytogenic Nomenclature 1985. The ideograms 25b and 25e in which the relative band positions and arm ratios were derived from actual
20 chromosome measurements is from Cytogenet Cell Genet 65:206-219 (1994).

 Figure 26 is an expression pattern of HU-Unc53/1 and HU-Unc532 in normal human tissues and cancer cell lines.

25 Figure 27 is a sequence map of Plasmid pNP3.

 Figure 28 is an exemplary list of prosite signatures which can be used to define and identify vertebrate homologues of UNC-53.

30 Figure 29 is a annotated sequence map of plasmid pEGFPsac. The GFP-C.elegans unc53sac fusion protein, and the C.elegans unc53 sac fragment are indicated.

 Figure 30 is a sequence map of plasmid pEGFP72. The GFP-C.elegans unc53 fusion protein and the C.elegans unc53 fragment are indicated.

35 Figure 31 is an annotated sequence map of plasmid

- 34 -

pEGFPsma. The GFP-C.elegans unc53sma fusion protein, and the C.e.unc53 sma fragment are indicated.

Figure 32 is an annotated sequence map of plasmid pEGFPec1. The GFP-C.elegans unc53ec1 fusion protein, and the C.elegans unc53 ec1 fragment are indicated.

Figure 33 is an annotated sequence map of plasmid pEGFPxba. The GFP-C.elegans unc53xba fusion protein, and the C.elegans unc53 xba fragment are indicated.

Figure 34 is an annotated sequence map of plasmid pLM4. Open reading frames of the hul-unc53/1 and GFP are indicated.

Figure 35 is a sequence map of plasmid pNP8.

Figure 36 is an illustration of microtubule association of C.elegans Unc53, shown in HepG2 cells, transiently transfected with pTB72, expressing C.elegans Unc53. panel A: microtubule staining of HepG2 cells, using YL1/2 panel B: C.elegans Unc53 staining, using rab4.

Figure 37 is an illustration of microtubule plus-end association in human cell lines transiently transfected with pTB72, expressing C.e.Unc53. C.elegans Unc53 was stained with mab-16-48. Panel C: COS cells showing microtubule association panel B: MCF7 cells showing microtubule plus-end association panel A: HepG2 cells showing microtubule plus-end association.

Figure 38 is an illustration of microtubule association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc53 fusion protein. GFP fluorescence was observed in living cells. Panel A: microtubule association of the GFP-C.elegans unc53 fusion protein panel B: microtubule plus-end association of the GFP-C.elegans unc53 fusion protein.

Figure 39 is an illustration of microtubule

- 35 -

association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc-53 fusion protein. Microtubules were stained with YL1/2 after paraformaldehyde fixation. Panel A: Microtubule association of the GFP-C.elegans unc53 fusion protein. Panel B: tubuline staining. Panel C: panel A plus panel B: co-localisation of the GFP-C.elegans unc-53 fusion protein and Tubuline can be seen as yellow.

Figure 40 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPsma, expressing the GFP-C.elegans unc53sma fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53sma fusion product. Panel B: Centriole association of GFP-C.elegans unc53sma fusion product when expressed at low levels.

Figure 41 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPec1, expressing the GFP-C.elegans unc53ec1 fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53ec1 fusion product. Panel B: Centriole association of GFP-C.elegans unc53ec1 fusion product when expressed at low levels.

Figure 42(a)/Figure 42(b) are illustrations of fluorescence of GFP in N4 cells transiently transfected with pEFPxba and pEFGPsac respectively.

Figure 43 is an illustration of microtubule association of in N4 cells transiently transfected with pLM4 expressing GFP-Hu-UNC53/1 fusion protein. Panel A: microtubule association of GFP-HU-UNC53/1 fusion protein. Panel B: microtubule plus-end association of GFP-Hu-UNC53/1 fusion protein. Panel C: microtubule association of GFP-Hu-UNC53/1 in dividing cells (end of division).

Figure 44 is an illustration of the sequence of Plasmid pNP9.

- 36 -

Figure 45 is an illustration of immuno fluorescence in melanoma G361 cells stained with sera 28.1. Panel A: Microtubule plus-end association of Hu-UNC53/1. Panel B: microtubule plus-end association of hu1-Unc53 in growth cone extensions.

Figure 46 is an illustration of GFP fluorescence and immunofluorescence in N4 cells transiently transfected with pLM4, and stained with sera 28.1. Panel A: Fluorescence of GFP-Hu-UNC53/1 fusion protein. Panel B: Immunofluorescence of serum 28.1.

Figure 47 is an overview of the microtubule (+) end, the microtubule and f-actin cytoskeleton binding properties of different constructs

Figure 50 is an illustration of rescue of lateral ALN neurons in mutant unc-53.

Dorsal view of the ALN neurones axones visualise in GFP fluorescence with the transgene pA/GFP in the posterior of an adult, (c) cellular body.

a) wild type, anterior axon (aa) migrates in a straight line along the body until reaching the head, on the dorsal sublateral cord, posterior axon (ap) migrates into the tail;

b) *unc-53(n152)*, anterior axons are the shorter, stop ahead of the vulva region and form numerous collateral branches towards the dorsal cord;

c) *unc-53(n152),pA/unc-53* anterior axons no longer form branches and migrate in a straight line into the head, as in the wild type at a).

scale bar 10 μ m.

Figure 51a : is an illustration of chimeric fusion between C. elegans and human 1 homologue of the unc-53 gene. The region of the putative nucleotide binding domain (NTP) is replaced in the C. elegans cDNA by the same region of the human homologue 1 of unc-53 (H1). The cDNA is under the promotor region A

- 37 -

(pA) of unc-53, which raise expression in the ALN lateral neurons.

Figure 51b : is an illustration of the chimeric minigene nematode/human pA/unc-53-H1 partially rescue the defect in the longitudinal migration of the lateral neurons ALN and PLN. The four strains compared are : wt; unc-53(n152); unc53(n152),pA/unc-53; unc-53(n152),pA/unc-53-H1. The observed phenotypes are put in three classes :

½sauvage₇, the axon is straight, unbranched, and migrates until the head; ½vulve₇, the axon is straight, unbranched, and stops in the vulva region; ½mutant₇, the axon is short, never joints the vulva region and made a lot of collateral branches. Numbers are in percentage. The number of observed axons are noted in the last column. The chimeric fusion between the C. elegans gene and human homolog (unc-53-H1) partially rescues the mutant phenotype. The chimeric gene was maded by replacing the putative nucleotide binding region (NTP)of the nematode cDNA by the same region of the human homolog 1 (H1).

Figure 52 is an illustration of the sequence for plasmid pLM5.

Figure 53 is an illustration of the sequence for plasmid pLM6.

Figure 54 is an illustration of the sequence for plasmid pLM1.

Figure 55 is a sequence map of plasmid pCB251.

Figure 56 is a sequence map of plasmid pNP10.

Figure 57 is a sequence map of plasmid pCB501.

Figure 58 is a sequence map of plasmid pTB115.

Figure 59 is a sequence map of plasmid pPD95.75.

Figure 60 is a sequence map of clone X16.

Figure 61 is a sequence map of plasmid pLM3

- 38 -

DEPOSITED MATERIALS

	Deposit	Date	Acc. Nr
	pCB201 plasmid DNA in E. coli	3 December 1996	LMBP 3594
5	Lambda clone 3B encoding hu-unc-53/1	3 December 1996	LMBP 3595
	MCF-7 clone z4 (mock)	3 December 1996	LMBP 1600CB
	MCF-7 clone (pCB201)	3 December 1996	LMBP 1601CB
	NIH-3T3 mock	3 December 1996	LMBP 1602CB
	NIH-3T3 pCB201	3 December 1996	LMBP 1603CB
10	pLM1	13 November 1997	LMBP 3762
	pLM4	13 November 1997	LMBP 3763
	pEGFP72	13 November 1997	LMBP 3764
	pCB501	13 November 1997	LMBP 3765
	BAC clone comprising fragment of hu-unc53/2 gene	15 November 1997	LMBP 3773
15	Worm strain with chimeric C.elegans/human unc-53 gene	15 November 1997	LMBP-1663CB

20

The above plasmids and cell lines were deposited at the Belgian Coordinated Collections of Microorganisms (BCCM) at laboratorium voor moleculaire biologie - plasmidencollective (LMBP) B9000, GENT, Belgium, in accordance with the provisions of the Budapest Treaty of 28 April 1977.

25

The present invention will now be described with reference to the following examples which are not limiting.

30

Identification of a human homologue of the UNC-53 protein of C.elegans.

Extensive searches with the ce-UNC-53 sequence (Figures 1 and 2) against the public domain databases

- 39 -

(EST, Genbank, EMBL, Swissprot and PIR) revealed no statistically significant homologies (a smallest sum probability (ssp) of $10 e - 8$ is generally accepted to be significant at amino acid level). Two ESTs

5 gbH09036 (ssp = $1.1 e - 5$) a Homo sapiens cDNA clone and gbAA049124 (ssp= $8.6-5$) a mouse cDNA clone showed homology to a "coiled coil" region a common motif in the contributing to protein secondary structure. (figure 3)

10 All other candidate scores were are at background level (ssp >0.21). Careful examination of weak candidate ESTs identified EST gb:R41071 from Homo sapiens, which had obtained a low score of 53 and a non-significant probability score of 0.33 (Fig. 4).

15 The inventors surprisingly discovered potentially significant homology with the Ce-UNC-53 nucleotide binding domain, provided multiple frameshifts and sequence errors were hypothesized.

The inventors amplified, cloned and sequenced
20 part of gb:R41071 from human heart and human lung cDNA and from human genomic DNA and discovered that clone gb:R41071 had up to ten 10 different mistakes in the region checked. 5 extra nucleotides were scattered along its sequence and two nucleotide substitutions
25 were identified, and gb:R41071 lacked three nucleotides present in our clone (Fig. 5). The novel sequence obtained was two nucleotides shorter and showed the two UNC-53-homologous regions in frame. The genomic fragment obtained is larger (700 bp total
30 length) than the corresponding cDNA clones indicating the presence of an interverting sequence of around 500 bp in nucleotide 162 of this fragment. The amplified cDNA fragment which was cloned to vector PCRII (Intvitrogen) and named pCR231 and was used as a probe
35 to screen cDNA libraries.

- 40 -

The conceptual translation of the clones we obtained by PCR were screened using blast and tblastn against all known protein and DNA sequences in the database. The only clone which came up with statistically significant similarity was Ce-UNC-53 (Fig.6). This human clone and Ce-UNC-53 thus form a unique homologous pair compared to the rest of the known sequences, indicating the statistical relevance and novelty of our discovery. We designate this human gene as hu-UNC-53/1. Human heart and a human colorectal adenocarcinoma cDNA libraries were probed with pCR231 probe to identify longer cDNA clones. The clones overlap giving a linear sequence of 3706 bp (Fig 8 and 26). This sequence shows an 959 amino acid open reading frame from the beginning of the clone. The absence of a 5' untranslated region suggests that the mRNA will extend 5'.

Sequence alignment searches of the public domain databases with the DNA sequence of hu-UNC-53/1 and its' conceptual translation identified a series of ESTs most of which correspond to the 5' UTR region. (Figures 7 and 8). Surprisingly, hu-UNC-53/1 identified also the cDNA clones gbH09036 and gbAA049124 homologous to the predicted coiled coil region in Ce-UNC-53 hu-UNC-53/1, and furthermore identified a third weakly homologous EST gbr21023. The inserts of gbH09036, gbAA049124 and gbr21023 were obtained from the Merck consortium and sequenced.

gbAA049124 is >95% identical to Hu-UNC-53/1 over 604 available amino acids (fig. 10) and is the mouse orthologue of Hu-UNC-53/1. The insert in gbH09036 is clearly homologous to hu-UNC-53/1 but derived from a different locus. We therefore name the gene identified by gbAA049124 Mu-UNC-53/1 and the gene identified by gbH09036 Hu-UNC-53/2. (Figure 11).

- 41 -

5 domains of high similarity mark the *unc-53* gene family

5 Ce-UNC-53 and the here-identified vertebrate
homologues form a unique novel protein family, that is
distant from the remainder of the proteins in the
public domain. Alignment of the predicted open
reading frames shows that Hu-UNC-53/1 and Hu-UNC-53/2
10 are equidistant from Ce-UNC-53. The highest homology
is found in the carboxyterminal amino acids of Ce-UNC-
53 region. The presence of a conserved GXXGKS/T box
suggests a nucleotide binding function. However, this
domain as a whole does not belong to a class of known
15 nucleotide binding proteins.

 The similarity amongst the presently known
sequence of the UNC-53 family of proteins is highest
in 5 blocks over most of the available sequence (959
amino-acids) and a further block identified in Figure
20 12a. These blocks can be assigned signature sequences
as displayed in figure 28 or can be assigned weight
matrices based on the alignment between the different
family members. By using truncated constructs of Ce-
unc-53, the functional relevance of these domains has
25 been addressed.

**Hu-UNC53/1 and Hu-UNC-53/2 are complex
transcription units.**

30 1. A cancer cell line RNA blots probed with HU-
Unc53/1.

 A Northern blot of poly-A+RNA from several
cancer cell lines (Melanoma G361, Lung Cancer A549,
Colorectal Adenocarcinoma SW480, Burkitt Lymphoma
35 DRajii, Leukemia Molt4, Lymphoblastic Leukemia K562,

- 42 -

HeLa S3 and Promyelocytic Leukemia HL60) was probed using the whole insert of pHH3b. No or weak expression was detected in the Burkitt Lymphoma DRajii, the Leukemia Molt4 and the Promyelocytic Leukemia HL60 cell lines. Five different transcripts are detected in the remaining cancer cell lines: transcripts 1 and 2 are larger than 9.5kb, transcripts 3 and 4 are 6 to 7 kb and the fifth transcript is around 6 kb. Transcripts 1 and 2 are present in all experssing cell lines. Transcripts 3 and 4 are restricted to Melanoma G361, Lung Cancer A549 and Colorectal Adenocarcinoma SW480 and are the predominant transcripts in Melanoma G361 and Colorectal Adenocarcinoma SW480. Transcript 5 is restricted to Lymphoblastic Leukemia K562 and HeLa S3 and is predominant in HeLa S3.

2. Cancer cell lines RNA blots probed with HU-UNC-53/2.

A similar set of cancer cell line Northern blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. HU-UNC-53/2 is expressed in Melanoma G361, Colorectal Adenocarcinoma SW480, Lymphoblastic Leukemia K562 and HeLa S3. No expression was detected in Lung Cancer A549, Burkitt Lymphoma DRajii, Leukemia Molt4 and promyelocytic leukemia HL60. Interestingly only 2 transcript sizes were detected of around 7 kb expressed in Lymphoblastic Leukemia K562 and HeLa S3 and a transcript of >9.5 kb in Melanoma G361 and Colorectal Adenocarcinoma SW480.

3. Normal Human tissue probed with HU-Unc53/1.

A Northern blot of poly-A+RNA from normal

human tissue was probed using the whole insert of phage HH3b. Expression levels are low in all tissues with the highest level in heart and placenta, several fold lower levels in brain and testis, even lower levels in skeletal muscle, pancreas, thymus, colon, small intestine, ovary and prostate. Expression in peripheral blood leukocyte, lung, liver, kidney, spleen is barely detectable.

4. Normal Human tissue probed with Hu-unc53/2.

A similar set of blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgtgagtc. Expression levels are low in all tissues with the highest level in kidney, lower levels in heart, placenta, lung, skeletal muscle and pancreas. Expression is barely detectable in brain and liver.

The hu-UNC53/1 and hu-UNC-53/2 homologues are clearly highly regulated genes, showing a strong tissue specificity and, probably, additional mechanisms of regulation (ie differential splicing of different promoters). The different proteins derived from RNA's identified by probe hh15 presumably share the carboxyterminal nucleotide binding domain. Ce-UNC-53 was shown to be a complex genetic locus and complex transcription unit. The different transcripts are thought to be a mechanism to assure the necessary specificity and functional diversity of this signal transduction pathway, with respect to different signals and receptors, different tissues and different directions of migration. The occurrence of a new transcript or the observed changes in expression levels in the cancer cell line blot suggests a role

for hu-UNC-53/1 and hu-UNC-53/2 in the establishment or maintenance of the transformed state of those cells.

5 **Phenotypic changes in cells transfected with the Nucleotide Binding Domain of Ce-UNC-53/1 and Hu-UNC-53/1**

10 Ectopic expression of full length Ce-UNC-53 in *C. elegans*, murine neuroblastoma cells or human MCF-7 breast-carcinoma cells, has been found to lead to increased filopodia outgrowth and increased motility (unpublished). The structure of Ce-UNC-53 protein is reminiscent of that of large kinases or dynamin where
15 a catalytic domain is positively or negatively regulated by domains that interface with signal transduction pathways for example (by GRB2 binding, phosphorylation or the like). The inventors therefore decided to test whether the nucleotide domain by
20 itself is capable of inducing the observed changes in the microfilament cytoskeleton and motile or ruffling behaviour.

 cDNA fragments coding for the nucleotide binding domains of Ce-UNC-53 and Hu-UNC-53/1 were cloned in
25 mammalian expression vectors with the CMV promoter (see experimental procedures).

 To be able to detect expression from pCB201 (Fig. 13), an N-terminal his and a T7 epitope tag were fused in frame with the hu-UNC-53/1 cDNA hh15. pCDU3
30 contains a larger fragment of Ce-UNC-53 and starts just before the conserved "VIELKIEL" domain (Fig. 12).

 The empty pCDNA3 vector or pCDNA3.1-His-LacZ, a mammalian expression vector for *E. coli* Beta-galactosidase, was used as a control vector (mock
35 transfection). The differences between mock and

- 45 -

transfected N4 and MCF-7 clones were analysed using phase-contrast and Nomarski microscopy coupled with time lapse analysis, phagokinesis and immunocytochemical characterisation of the F-actin.

5

Phenotypic changes in mouse N4 neuroblastoma cells

10 N4 neuroblastoma cells were stably transfected with control construct pCDNA3.1 and the C. elegans UNC-53 constructs pTB72, pCDU2, pCDU3 and pCDU4. The population of clones transfected with the empty expression vector were homogeneous and similar to wild type N4 cells. In contrast thereto, 1/4 to 50% of the clones transfected with pTB72, pCDU2, pCDU3 and pCDU4 (see experimental procedures and Figs. 1,17,15 and 16 respectively) had distinct phenotypes:

1. Wild type or N4 cells transfected with pCDNA3, designated as mock transfection show a central cell body, with extensions, designated as neurite outgrowths. Less than 5% of the population have lamellae. When present, they are generally situated on the cell body and on the opposite site of the neurite extensions (figure 22a). The lamellae show a radial actin spike pattern. Limited branching of the actin fibres is observed in wild type or pCDNA3 transfected N4 cells. Side branches are smaller and can be clearly distinguished from the main actin branch (figure 23a).

2. N4 cells, stably transfected with pCDU4, harbouring the homology block E, show an overall morphology which is similar to that of wild type N4's (a cell body with neurite outgrowth). They exhibit however an increased frequency and level of lamellae formation (figure 22b). These lamellae, which contain

- 46 -

F-actin microspikes are found on both the cell body and the neurite outgrowth (figure 23b). Wild type N4 cells, in contrast thereto, rarely exhibit lamellae on the neurite outgrowths.

5 3. N4 cells, stably tranfected with pCDU3, encoding for homology blocks C, D and E, show an even higher level of lamellae formation labelled with TRITC-phalloidin, the cells appear surrounded with F-actin fibres, consisting of bundles of F-actin
10 microspikes (figure 23c). The presence of these lamellae has completely modified the general appearance of the cells. They appear flatter and in 90% of the population, it is not possible to distinguish between the cell body and the wide neurite
15 as they flow gradually into one another (figure 22c). If wild-type-like thin neurite-like outgrowths are present, they are frequently numerous, branched and located all around the cell.

20 4. The overall morphology of N4 cells, stably transfected with pCDU2, encoding for homology blocks A, B, C, D, and E, resembles that of the wild type cells since, cell body and neurite outgrowth can be clearly distinguished. The pCDU2 transfected cells however show more neurite outgrowth, and these are
25 long and very branched, especially at the end of the outgrowth. When neurite outgrowths of different cells make contact, increased branching can be observed, giving the appearance of a network (figure 22d). N4 cells, transfected with pCDU2, show bundles of long
30 radial F-actin filaments (microspikes), which can be branched, especially apically. The space between the hand-shaped actin spikes is mostly filled in with actin, leading to small lamellae-like structures. Also the network-like branching between the cells
35 shows both the bundled actin structures and the

- 47 -

lamellae-like fill-in features. These dense F-actin structures are sometimes seen on the cell body, which enhances the network-like appearance of the cells (figure 23d).

5 5. N4 cells, stably transfected with plasmid pTB72, encoding the full length C. elegans UNC53 protein, seem to have a more rigid structure than wild type cells, most clearly seen as spindle-like and triangle-like cells. The corners of these cells show
10 an increased level of hand-like lamellae structures. This specific phenotype is best seen when the cells are grown at low density (figure 22e, Fig. 23e).

15 Phenotypic changes in human breast carcinoma MCF-7 cells

MCF-7 cells were stably transfected with the pTB72, pCDU2, pCDU3, pCDU4 and pCB201. The population
20 of clones transfected with the LacZ-expression vector were homogeneous and similar to wild type MCF-7 cells. In contrast thereto, ~30-50% of the clones transfected with pTB72, pCDU2, pCDU3, pCDU4 and pCB201 had distinct phenotypes which were analysed as above for
25 the N4 cells:

1. Wild type and mock (pcDNA3) transfected MCF-7 cells are heteromorph. In general they are round cells or clusters of cells surrounded by lamellae. Bulges, similar to thick filopodia, can be observed
30 (figure 19a). When the cells are stained with FITC- or TRITC coupled phalloidin, F-actin stress fibres can be observed, often in rings surrounding the cell body (figure 20a & 21a). When cells are round up like this actin is present at the edge of the cell
35 body. Less than 10% of the cells display filopodia

- 48 -

filled with radial F-actin microspikes. In time-lapse analysis the cells are highly quiescent with limited ruffling at the edge of the cell.

2. MCF-7 cells transfected with pCDU4, encoding
5 for homology block E, show two major phenotypic
differences compared to the wild type cells. These
cells are more flat and have more extended
lamellipodia leading to a pancake-like appearance.
Some clones show more filopodia than wild type (figure
10 19b). Radially organised F-actin fibres can clearly
be observed in the lamellae surrounding the cells.
These stress fibres resemble the wild-type structures,
but have a more radial than circular orientation. In
the filopodia, one can observe an increase of
15 apparently unorganised, bundles of actin patches
(figure 21b).

3. MCF-7 cells, stably transfected with pCDU3,
encoding the homology blocks C, D, and E, shows a
strikingly different and constant morphology. The
20 cells appear smaller than wild type because they are
more rounded up. All the cells have more filopodia,
surrounding the cell body (figure 19c).
Morphologically these filopodia have the same "hand-
like" appearance as those observed in N4 neuroblastoma
25 cells. Such filopodia are hardly ever observed in
mock transfected MCF-7 cells. These filopodia are
filled with F-actin fibres. Compared to wild type
cells, fine actin stress fibres are decreased (figure
21c). In time-lapse analysis single cells as well as
30 clusters of cells can be seen to ruffle much more
dynamically than single or clusters of wild type
cells. The "half-life" of a filopodia outgrowth on
the cell surface is much shorter in transfected cells
and the numbers of filopodia present at any time
35 higher.

- 49 -

4. Cells transfected with pCB201 (which is structurally similar to pCDU4 but human) has a phenotype that is nearly indistinguishable from that of cells transfected with pCDU3 except that the observed phenotype and ruffling activity and filopodia outgrowth is even higher than pCDU3 (figure 18).

5. The overall morphology of the MCF-7 cells transfected with pCDU2, which encodes the homology blocks A, B, C, D and E, resembles that of the pCDU3 transfected cells. The cells are more rounded up and show more filopodia than the wild type and mock transfected cells (figure 19d). The filopodia, which are all around the cell body tend to be longer, and show a difference in actin organisation. The small filopodia have the same actin bundles as seen in the pCDU3 transfected cells. In the longer filopodia, the actin bundles are more parallel, and radial to the cell body (figure 21d).

6. MCF-7 cells transfected stably with pTB72, encoding the full length UNC53 protein, are extremely rounded up, and tend to adhere more than wild type cells. The cells grow in clusters with sausage- or tube-like shapes. The presence of large extremely thin lamellae with a surface area of more than three times the central cell body forms a second morphological feature, unique for the pTB72 transfected MCF-7 cells (figure 19e). These sheets are difficult to observe under a phase contrast microscope, but are very clear when stained with phalloidin. The lamellae protrude from one side of a cell or group of cells and are filled with thin long criss-crossing actin fibres, different from "giant" wild type MCF-7 cells (figure 21e).

These experiments lead to the following set of conclusions: (Figure 47 summarises the data of the

- 50 -

domain swapping experiments in *C. elegans unc-53*)

1. Murine and human cells transfected with the Ce-UNC-53 or hu-UNC-53/1 domains show clear effects on the nature and dynamics of their motile behaviour as demonstrated by changes in the F-actin cytoskeleton (the increase in lamellipodia, hand-like filopodia and "hair-like" microspikes on the cell surface and the associated reduction of the "rings of F-actin" stress-fibres).

2. This effect is found in two cell types of different species and tissue origin: MCF-7 cells (human breast carcinoma cells of epithelial origin) and murine N4 neuroblastoma cells. pCB201, pCDU3 and pCDU4 induce in MCF-7 cells a type of filopodium which is frequent in wild type N4 cells but rare to absent in wild type MCF-7 cells, suggesting the activation by these constructs of motile behaviour which is "normal" in N4 cells but of an unusual type in MCF-7 cells. This indicates the activation of a specific downstream process as opposed to a disruption of an existing process. It is well known that some cell types prefer to migrate with filopodia and other cell types with lamellipodia.

3. Expression of pCB201, pCDU3 and pCDU4 gives qualitatively similar F-actin remodelling and increased filopodia and lamellipodia outgrowth. pCB201 and pCDU3 are however much more active in this process than pCDU4.

4. pCB201 is a much more potent activator of filopodia outgrowth than pCDU4, which is to be expected considering the large evolutionary distance between *C. elegans* and vertebrates.

5. These experiments identify homology domain E (predicted nucleotide binding domain) of UNC-53 as the

- 51 -

"domain" that activates F-actin remodelling and filopodia/lamellipodia outgrowth. Progressive addition of the aminoterminal homology A,B,C,D lead to qualitative and quantitative modulation of the phenotype present in domain E.

6. Homology domains C and D (pCDU3) "enhance the basic activity present in homology domain E (pCDU4/pCB201).

7. Homology domains B and C (pCDU2) qualitatively modify the phenotype of domain E, leading morphologically different lamellipodia formation than pCDU3 transfected cells. It is thought that lamellipodia and filopodia formation are mediated by different signal transduction pathways requiring two related but different Ras-like G-proteins RAC for lamellipodia formation and CDC42 for filopodia formation.

8. pTB72 which includes homology domains A,B,C,D,E plus an additional 700 amino acids not yet identified isolated in the human members of the family confers a more localised filopodia outgrowth and a different morphology.

9. The expression levels of pTB72 (full length C. elegans UNC-53), pCDU3, pCDU4 and pCB201 are extremely low. The observed effect is therefore unlikely to be due to dominant negative effects (such as stoichiometric depletion of other cellular components) or structural changes in the actin cytoskeleton mediated by UNC-53 or its fragments.

The data point to a multi-domain organisation in UNC-53 whereby the aminoterminal domains exert positive (e.g. pCDU3) and negative (e.g. pCDU2) control on the activity of the domain E or are leading to novel activities or the localiation of the activity in the cell (pCDU2, pTB72). Our observation that the

- 52 -

nucleotide binding domains (NTB) of distantly related members of the UNC-53 family induce similar phenotypes, suggests a general role for this domain of the UNC-53 family.

5

CELLULAR ASSAYS TO IDENTIFY PHARMACOLOGICAL
MODULATORS OF UNC-53 AND COMPONENTS OF THE UNC-53
PATHWAY

10

Mammalian and human cells transfected with plasmid constructs containing unc-53 sequence of either C. elegans or of human origin were observed to display obvious, specific and similar changes in comparison to mock or untransfected parent cells. These changes relate to the functioning of the cytoskeleton, in particular the F-actin cytoskeleton, to cell locomotion and directionally cell motility and reflect UNC-53 gene family members as capable of playing an integrator function in cell motility.

15

20

The cellular tools derived through transfection and derived functional assays with these cells not only enable characterisation of the motile phenotype typically observed after introduction of unc-53 genes, they also can be easily adapted to screen for pharmacological compounds that interfere with either (1) the expression of unc-53 gene family members, (2) the cellular functioning of unc-53 transgene(s) and of components in the unc-53 signal transduction pathway.

25

30

Two classes of pharmacological modulators are envisaged.

A first class are inhibitors of UNC-53s or the unc-53 pathway(s), which revert the described phenotypic changes induced by unc-53 transgenes or

35

aspects thereof. Such compounds are considered relevant leads to target diseases where unwanted directional motility of cells occurs such as metastasis, angiogenesis or inflammation.

5 Secondly, pharmacological stimulators are envisaged, such as compounds which induce - in non-transfected cells - phenotypes that induce or mimic (aspects of) the described 'unc-53' phenotype. Such compounds may do so by inducing or upregulating
10 expression levels of a known unc-53 gene or by activating endogenous (yet unidentified) members of the unc-53 gene family. The target application here are wound and tissue repair, in particular diseases such as neuronal regeneration and plasticity.

15 The nature of compounds envisaged can be small (organic) molecules, bio-molecules (such as peptides, sense or antisense (oligo-)nucleotides or chemical modifications thereof. Alternatively, compounds can be thought of as a series of plasmid nucleotide
20 constructs containing gene sequences in a screen for novel unc-53-unrelated genes with a similar functional effect in the cell or genes related to the unc-53 gene family or novel members of the unc-53 gene family based on sequence similarity such as for example the
25 genes in plasmids pTB72, pcDU3, pcDU4, pcDU2, pcB201, or modifications thereof such as for example epitope tagged, deletion, complementation or mutagenised nucleotide constructs.

The cellular assays envisaged in the claims have
30 been exemplified for three cell lines: the human breast carcinoma cell line MCF-7, the mouse neuronal cell line N4 and the mouse fibroblast cell line NIH-3T3. Pharmacological assays are focused on quantification of endpoints in a high throughout
35 screening mode. Many of the computer aids for

- 54 -

(semi-) automation are well known to the field and currently applied in the applicants labs. Given the subtlety of the phenotypes observed, primary focus was given to morphological assays that assess the phenotypes or aspects thereof.

The nucleotide binding domain of Hu-UNC-53/1 has transforming activity in NIH3T3 fibroblasts

Biochemical and genetic analysis suggest that UNC-53 functions in GRB-2 mediated signal transduction pathways controlling cell motility. The occurrence of an altered hu-UNC53/1 mRNA pattern in cancer cell lines, moved us to investigate if whether hu-UNC53/1 plays a role in the transformed state of those cells.

Thereto, we tested the ability of the nucleotide binding domain of hu-UNC-53/1 and Ce-UNC-53 to transform NIH/3T3 cells. Construct pCB201 (hu-UNC-53), which induces ruffling behaviour and cell motility, were transfected into NIH3T3 cells. Positive controls included Myc and H-ras. Negative controls included empty vector and Rac 1N17 and cdc42N17.

The cells that survived G418 selection were assayed for loss of contact inhibition (their ability to grow as foci). Positive controls included the combination of two well known oncogenes Myc and H-ras which were able to produce a high number of foci. The nucleotide binding domains of both Ce-UNC-53 and hu-UNC-53/1 are able to induce foci in this assay (Fig 24 & Table 1).

35

Table 1 Foci formation in NIH-3T3 cells stably transfected with pcB201

	mock	pcB201
5	22	138
	59	143

10

This suggests that the function of UNC-53 is not restricted to the activation of motility. UNC-53 may exert this additional function through the activation of as yet to be identified signal transduction pathways. Oncogenes frequently arise when a "controlling" domain and "activation" domain are separated through chromosomal rearrangements or integration of a part of a gene in the oncogenic virus. E.g. Erb Receptor tyrosine kinases, Ost a nucleotide exchange factor for Rac-1.

20

Hu-UNC-53/1 is localized to chromosome 1q31.1

Clone F226 (BACH-135 (014), Genome Systems, inc) was isolated from a human genomic BAC library using PCR231 as a probe and was confirmed by sequence analysis to be derived from the hu-UNC-53/1 locus. Purified DNA from clone F226 was labeled with digoxigenin dUTP by nick translation. Labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextran sulfate and 2X SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluoresceinated antidigoxigenin antibodies followed by counterstaining with DAPI. The initial experiment resulted in

25

30

35

- 56 -

specific labeling of the long arm of a group A chromosome. A second experiment was conducted in which an anonymous probe which was previously mapped to lp34 and confirmed by cohybridization with a chromosome 1 centromere specific probe, was cohybridized with F226. The experiment resulted in the specific labeling of the long and short arms of chromosome 1. Measures of 10 specifically hybridized chromosomes 1 demonstrated that F226 is located at a position which is 52% of the distance from the heterochromatic-euchromatic boundary to the telomere of chromosome arm 1q, and that corresponds to band 1q31. At total of 80 metaphase cells were analyzed with 72 exhibiting specific labeling (Fig. 25).

Gains of DNA sequences in 1Q31 were found in more than 10% of primary bladder tumors (Genes Chromosom Cancer 12: 213-219 (1991)). A putative tumor suppressor gene located near the locus F13B on chromosome arm 1q31-q32 appears to be involved in the pathogenesis of medulloblastoma (Int. J. Cancer 67: 11-15 (1996)). Loss of heterozygosity in this region of chromosome I has been implicated in development of human hepatoblastoma. Partial trisomies of 1q31 were found in Ewing's Sarcoma cell lines isolated from patients Cancer Genet Cytogenet 12: 1-19 (1984).

HU-UNC-53/2 is localised to Chromosome 11p15.1

DNA from clone F329 from BAC for Hu-unc-53/2 was labeled with digoxigenin dUTP by nick translation and applied in the experimental settings used for FISH of Hu-unc53/1 with F226. The initial experiment with F329 resulted in the specific labeling of the mid short arm of a group C chromosome which was believed to be chromosome 11 on the base of size, morphology

- 57 -

and banding pattern. A second experiment was conducted in which a biotin labeled probe specific for the centromere of chromosome 11 (D11Z1) was cohybridised with clone F329. This experiment
5 resulted in the specific labeling of the centromere in red and the mid short arm in green of chromosome 11. Measurements of 10 specifically labeled chromosomes 11 demonstrated that F329 is located at a position which is 65% of the distance from the centromere to the
10 telomere of the chromosome 11p, an area which corresponds to band 11p15.1. A total of 80 metaphase cells were analysed with 72 exhibiting specific labeling.

15 Chromosome 11p15 is a region showing loss of heterozygosity (LOH) in a variety of human malignancies, primarily breast cancer (Ali et al., Science 238, 185-188 (1987); Wingvist et al., Cancer Res. 53, 4486-4488 (1993)) but also Wilms' tumor
20 (Dowdy et al., Science 254, 293-295 (1991); Cowell et al., Br.J.Cancer 67, 1259-1261 (1993)), ovarian and testicular malignancies (Lothe et al., Genes Chromosomes Cancer 7, 96-101 (1993); Weitzel et al., Gynecol Oncol. 55, 245-252 (1994)) stomach cancer
25 (Baffa et al., Cancer Res. 56, 268-272 (1996)), lung cancer (Ludwig et al., Int.J.Cancer 49, 661-665 (1991); Fong et al., Genes Chromosomes Cancer (1994)), infantile tumors of adrenal and liver (Byrne et al., Genes Chromosomes Cancer 8, 104-111 (1993)). Since
30 LOH is believed to indicate inactivation of a tumor suppressor gene at the location where LOH occurs, the frequent LOH found at 11p15 in multiple human cancers suggests the presence of either a cluster of tumor suppressor genes or a single tumor suppressor in this
35 region (Seizinger et al., Cytogenet. Cell genet. 58,

- 58 -

10080-10096 (1991)). Chromosome transfer studies have shown that chromosome 11 can suppress tumorigenicity of both human breast cancer (Negrini et al., Cancer Res. 55, 3003-3007 (1995)) and Wilms' tumor cells (Dowdy et al., Science 254, 293-295 (1991)) and a gene (named HTS1 or ST5) that may be responsible for suppressing tumorigenicity in HeLa cells has been mapped to 11p15 (Lichy et al., Cell Growth Diff. 3, 541--548 (1992)). Abnormalities at 11p15 have also been identified in a variety of other cancers, including lung cancer (parental origin of 11p15 deletion) (Kondo et al., Oncogene 9, 3063-3065 (1994)), bladder cancer (Presti et al., Cancer Res. 51, 5405-5409 (1991)), myeloid leukemia (translocation) (Nakamura et al., Nat. Genet. 12, 154-158 (1996)), malignant astrocytomas and other primitive neuroectodermal tumors (deletions) (Fults et al., Genomics 14, 799-801 (1992)), rhabdomyosarcoma (Scrabble et al., Nature 329, 645-647 (1987)) and hepatocellular carcinoma (Fujimori et al., Cancer Res. 51, 89-93 (1991); Wang et al., Cell Genet. 48, 72-78 (1988)). Recently a gene, TSG101, was cloned that is mutated in human breast cancer and deleted in uncultured primary human breast carcinomas (Li et al., Cell 88, 143-154 (1997)).

DIAGNOSTIC ASSAY USING THE DNA SEQUENCE OF HUMAN UNC-53S

The differential expression of human unc-53 transcripts in Northern blots of normal tissues versus transformed cell lines and the chromosomal locus of hu-unc-53/1 at 1q31 being a locus linked to three diseases, suggests the potential implication of hu-unc-53 genes in oncogenesis. By using the complete

- 59 -

DNA sequence of hu-unc-53/1 or /2 or fragments thereof in FISH, the potential involvement of these genes can be diagnosed in patients as exemplified in figure 26. Alike, the use of these hu-unc-53 sequences in diagnostic PCR assays can be used to determine overexpression of hu-unc-53s or fragments thereof.

Assay for microscopic phenotypic UNC-53 transfected MCF-7 cells

Mock and unc-53 transfected MCF-7 cells were seeded at low density in culture plates and allowed to adhere to the vessel. Light microscopic inspection at different time points either on live cells or after chemical fixation with Karnovsky's fixative revealed that in pcB201, MCF-7 transfected cultures a rounded shaped cell body with at their boundaries many filopodia. In contrast, mock or untransfected clones had a predominant 'flat' phenotype - with little or no filopodia. Quantitative measurements confirmed the statistical significance of this shift in phenotype (table 2 below).

TABLE 2

Quantification of phenotypic changes in unc-53 transfected MCF-7 cells (*)

Transfection:	clone	no feet (**)	with feet (**)	fraction with feet
mock	e	34	8	0.19
		37	0	0
pcB201	2	17	92	0.84
		37	83	0.69
	16	27	62	0.70
		20	71	0.78
		13	85	0.87

(*) Clones were passaged thrice, frozen and stored.

- 60 -

Thawed cells were trypsinised at confluency, monodispersed, seeded in flasks and allowed to attach to substrate overnight to 48 hours. Cultures were fixed with Karnovsky fixative and inspected using phase contrast microscopy. In parallel experiments, resistance to genitacin was confirmed.

(**) values are expressed as cells per microscopic view.

10 Assay for ruffling and motile behaviour using
 automated time lapse

15 The dynamic changes in cells are well known in the field. Animations of e.g. actin ruffles in astrocytoma cells or of actin based cell motility in e.g. fibroblasts can be accessed
(<http://www.stc.cmu.edu/CLMIBhp/Imggallpg/Moviespg/actinruffle.mov>) or
(<http://util.ucsf.edu/mitchi/Movies/migration.html>) on
20 the world wide web. The dynamic changes as a result of transfection with unc-53 can best be appreciated in time lapse video sequences. At high magnification, the 'filopodia' display arrays of microspikes with highly dynamic behaviour. A rough visual estimate
25 suggests these phenomena to be at least 10-fold increased in pcB201 transfected cells relative to the mock-transfected MCF-7 cells. Animations of these clones in NIH-Image can be requested from author or applicant.

30 Time lapse video imaging probably is the most informative way to appreciate the unc-53-induced phenotype in MCF-7 and is amenable to high throughput screening in a pharmacological context. Time lapses compressing 5 minutes real time supply sufficient
35 information to quantitate the intensity of the motile

- 61 -

behaviour of pcB201 transfected MCF-7 cells in e.g. 12 well plates. In addition, algorithms have been described in the field which can automatically compute the 'motile area' of cells by comparing cells in two images appropriately spaced in time (van laerebeke et al., 1992, cytometry, 13, 1-8).

Assay for visualising unc-53-induced F-actin recruitment in MCF-7 cells

10

Cultures were chemically fixed, detergent extracted and fluorescently stained for F-actin (filamentous-actin) using fluorescently labeled phalloidin (Wieland et al., 1985, Int. J. Peptide & protein Res, 21, 3-10) which display in a more specific way the dramatic phenotypic changes to transfection with unc-53 transgenes. By using image capturing and analysis of the F-actin patterns, image analysis algorithms well known in the field can assess in an automated way, the f-actin filament positions, texture and distribution relative to the nuclear position or gravity point of the cells. Such algorithms are capable of discriminating phenotypic changes and thus also effects of pharmacological inhibitors of transgene-induced phenotypes as well as compound induced unc-53 like phenotypes in mock or untransfected cells.

15

20

25

30

Phagokinesis assay for unc-53-induced directionality and quantity of motility

The methods are described in the experimental section. Two cell populations with different motile behaviour in phagokinesis assays were observed. In table 3 below the fraction of mock and UNC-53

35

- 62 -

transfected MCF-7 cells that produced linear tracks in the phagokinesis assay are shown. In the mock transfected MCF-7 cells, 61% of the cells produce a round track (long and short axis less than 2-fold different) and 39% cells produced 'linear' tracks (long and short axis more than 2-fold different). pcB201 transfected MCF-7 cells produced an increase of the fraction of cells displaying 'linear' tracks to 50%. An increase in the fraction linear tracks was made for MCF-7 cells transfected with full sequence Ce-unc-53.

In addition, a significant increase of 50% in the median area of tracks of a culture vessel was observed in the pcB201 transfected MCF-7 cells versus mock transfected MCF-7 cells (Table 2). These observations suggest that pcB201 as well as pTB72 transfection into MCF-7 cells is capable of increasing *in situ* locomotion in Ce-UNC-53 MCF-7, e.g. by increasing spreading, ruffling, or other forms of non-directional motility in the 'round' population. In addition the Ce-UNC-53 transgene in MCF-7 cells drives a fraction of the MCF-7 cells from non-directional motility (round tracks) into directional migration (linear tracks). Clone 2 thus provides a tool to analyse inhibitory or stimulating effects of pharmacological compounds on directionality or quantity of cell motility in relation to UNC-53.

30

35

Table 3. Analysis of motility in phagokinesis assays

Track morphology: fraction linear tracks

5	plasmid	clone	round	linear	l/r
	Mock	Z4	18	13	0.42
			17	11	0.39
10			22	12	0.35
	pCB201	Clone 2	16	9	0.36
			13	13	0.5
			7	8	0.53
15			9	9	0.5
<i>Track Size</i>					
	Clone	average \pm SD	min	max	(N)
20	Z4	1626 \pm 188	1444	2011	(8)
	Clone 2	2326 \pm 283	1989	2816	(8)

25 **Assays for the localisation of unc-53 in the cell
to microtubules or microtubule (+) plus ends**

UNC-53s have been shown to reside on microtubules and preferentially on the microtubule (+)-ends of cells. This localisation represents an important feature of the UNC-53 family of proteins, which is rarely observed in other proteins. Absence of microtubule (+)-end binding in the protein APC following mutation has been implied in the role of APC in colon cancer (Smith et al., 1994, Cancer Res., 54, 3672). In analogy, it can be postulated that the proper functioning of UNC-53 also may depend on its specific localisation in the cell.

The methods used in the examples which prove the co-localisation with microtubules form a base for a series of assays for compounds which specifically

- 64 -

affect microtubule (+)-end binding of UNC-53s. To the skilled eye, the typical localisation of an UNC-53 protein on microtubules can be readily recognised and thus is sufficient for the interpretation of whether the treatment with a compound has affected the localisation of this UNC-53 (or a fragment thereof). Moreover, by combining the described methods (co-localisation) - well known to one skilled in the field and exemplified by the methods in the "experimental procedures" section - one can unequivocally confirm a compounds ability of abrogating (or promoting) microtubule and microtubule (+)-end binding.

Such an assay comprises contacting a cell culture of a cell line expressing an UNC-53 with a compound in the culture conditions proper for the said cell line, followed by an incubation and finally observation of the UNC-53 (or fragment) in situ by e.g. fluorescence microscopy (for GFP-chimeras) or by fixing the cell culture and performing an immunocytochemical staining for the UNC-53 (or fragment). For the co-localisation, methods such as immunocytochemistry for the microtubules of a cell or cell line combined with either immunocytochemistry for Ce-UNC-53 or Hu-UNC-53s or fluorescent detection GFP-UNC-53 chimeras are performed consecutively.

C.elegans-UNC-53 preferentially binds microtubule plus-ends or GTP-tubulin

Biochemical characterisation of UNC-53 has shown that UNC-53 binds the SH3 binding domains of SEM-5/GRB-2 and binds F-actin *in vitro*. GRB2 has been localised to the cortex of the cell and reported to be involved in the control of cell motility. To determine the *in vivo* subcellular localisation of Ce-

- 65 -

UNC-53, we transiently transfected COS, HepG2 and MCF7 cells with pTB72, an expression construct containing the full length Ce-unc-53 cDNA. This construct was previously shown to activate cell motility in N4 neuroblastoma and MCF7 cells. This construct gives high transient expression in COS cells, high to medium levels of expression in MCF7 cells and medium to low levels of expression in HepG2 cells. To visualise UNC-53, tubulin and F-actin, transfected cells were stained with various combinations of the anti-Ce-UNC-53 mab 16-48-2, rabbit anti-UNC-53 polyclonal, anti-tubulin mab YL1/2 and fluorescently labelled phalloidin.

At high levels of expression UNC-53 co-localises with the entire microtubule cytoskeleton, but at lower expression levels UNC-53 signal is restricted to the terminal regions of the microtubules at the plus ends. Very low levels of the expression yield a dot-like pattern in the vicinity of the cortex of the cell.

To map the MTB plus end domain of Ce-UNC53, we made two constructs pcDU2 (figure 17) and pcDU3 (figure 15) in which the aminotermus of Ce-UNC-53 is deleted. Proteins corresponding to these constructs are thought to be made *in vivo* from different unc-53 promoters. Transient transfections followed by immunolocalisation showed these proteins to be cytoplasmic. In stable transfections in N4 neuroblastoma cells and MCF7 cells they were shown to be no longer toxic to cells but cause highly increased activation of filopodia formation. We thus uncoupled (1) toxicity of Ce-UNC-53 from activation of motility and (2) microtubule binding from the activation of motility.

35

- 66 -

Analysis of the microtubule association of the
C.elegans and Human 1 UNC53

5 To isolate the microtubule association domain of
the C.elegans UNC53, N-terminal GFP fusions were made.
C-terminal deletions on the fusion product revealed
that the microtubule association was localised in the
N-terminal half of the protein. A GFP fusion was also
constructed with the Human1-UNC-53, to analyse the
10 microtubule association properties of this protein.
The association with microtubules was confirmed. A
mouse anti sera was used to show the presence of
native Unc-53 on microtubule plus ends of melanoma
line G361. The epitope recognition of the antibody
15 was confirmed by immunohistology experiments with
mammalian cells, transiently expressed with pLM4,
expression the GFP-hu1-UNC53 fusion protein.

Results

20

1. When transiently transfecting pTB72 in
several cell lines C.elegans UNC-53 associates with
microtubules and preferentially the plus-ends of the
tubuline fibres. Transfection of plasmids pCDU3 and
25 pCDU2 in N4 and MCF7 cell lines did not result in the
observation of microtubule co-localisation. pCDU4
resulted in no staining using mab 16-48 antibody (LMBP
Accession No. 1383CB) concluding that the epitope for
this antibody is localised outside the fragment
30 expressed by pCDU4.

It is possible that the microtubule associated
domain is situated in the N-terminus of the protein.
For this reason, we constructed an N-terminal GFP
fusion with the full length C.elegans UNC-53 sequence,
35 and various C-terminal deletion derivatives. These

- 67 -

fragments encode the N-terminal part of UNC-53 from 139 to 760 aa.

Furthermore, to analyse if the cloned fragment of hul-unc53 also could be associated with microtubules, a plasmid encoding a GFP fusion with the hul-Unc53 protein was constructed, and introduced into mammalian cells. A derivative of this construct was also constructed.

10 2.

a) Transient expression of C.elegans Unc-53 GFP fusion in N4 neuroblastoma lines

15 N4 cells where transiently transfected with pEGFP72, encoding a fusion protein of GFP and full length C.elegans unc-53 sequence. On an inverted microscope, the fluorescence of the GFP molecule could be followed in living cells. Cells which expressed low to medium levels of the fusion molecule showed a normal morphology after 18h to 30h. In these cells the co-localisation of the GFP fusion protein with the microtubules could clearly be demonstrated (figure 38a). In cells which demonstrated a low but still distinct GFP fluorescence, specific microtubule plus-end association could be observed (figure 38b). Cells expressing high levels of the GFP fusion protein tend to round up, in such a way that the microfilaments are difficult to visualise. After 48h, almost no GFP expressing cells can be found. It has previously been observed in transient expression of Unc-53, using plasmid pTB72, that the protein is toxic for the cells. The transient transfection experiments with the pEGFP72 plasmid gives the same observation, indicating that at least two features of the Unc53 protein are conserved in the GFP fusion protein, being

20
25
30
35

- 68 -

the microtubule association and the toxicity of the protein.

The transfected cells were fixed with paraformaldehyde, and the tubuline was stained using antibody YL1/2 and antimouse-CY3 (Jackson Labs). Although a significant loss of GFP fluorescence was observed, one could clearly demonstrate that the filaments observed with the GFP fluorescence co-localise with the microtubules staining (figure 39).

Putative Assay

Mammalian cells, in this case N4, were transfected with a lipofecting agent (lipofectAMINE) while in suspension, not being attached to a surface. After transfecting those cells with pEGFP72, the transfected cell suspension could be diluted in 24- and/or 96-well plates, enabling them to attach to the surface. Each well may contain a different compound of the collection to screen. After 24h, plates could be automatically screened for fluorescence levels. Wells containing a compound that abolish the toxicity of the GFP-C.elegans UNC-53 fusion protein will give high levels of fluorescence. Compounds having no effect on the fusion product will give no or only low levels of fluorescence.

b) Transient expression of the truncated GFP-C.elegans UNC-53 fusion proteins.

To assay if the microtubuline association did occur in the N-terminal part of the C.elegans Unc-53 protein, various C-terminal deletions were constructed.

Transfection of pEGFPsma and pEGFPec1 coding

- 69 -

for 760 AA and 670 of the N-terminal part of C.elegans
UNC-53 in fusion with GFP, resulted in microtubuline
association, as could be visualised in living cells.
The association with the microtubules is less abundant
5 than observed when expressing the full length
C.elegans UNC-53 protein, but fibres could clearly be
observed (figures 40a and 41a). More background
fluorescence is seen. This could be due to a lesser
association to the microtubules or to a instability of
10 the fusion protein. The association with microtubules
could not be observed after fixing the cells with
paraformaldehyde nor with methanol fixation, giving an
extra indication for the weak association with the
microtubule network of these proteins or potential
15 instability of the fusion protein. At low expression
levels the association of the GFP fusion protein with
the centrosomes could clearly be detected (Figures 40b
and 41b). Centrosomes are the location in the cell
with the highest microtubule concentration.

20

No plus-end associations could be observed
with the deletion constructs, even when cells where
expressing low levels of the GFP fusion proteins. In
the case of very low expressions, the centrosomes
25 could clearly be detected.

When transfecting N4 cells with pEGFPsac or
pEFPXba, coding for 139 aa and 256 aa of the N-
terminal part of C.elegans UNC-53 in fusion with GFP,
30 no microtubule association could be observed. This
indicates that at least 670 aa of the N-terminus of
the C.elegans UNC-53 is needed to have microtubule
association (figures 42a and 42b).

35

c) Transient expression of the GFP-hu-UNC-53/1

- 70 -

fusion proteins and a deletion derivative.

5 Plasmid pLM4 was transiently transfected into N4 neuroblastoma cells, and GFP fluorescence was observed in living cells. GFP fluorescence of the available sequence of hul-UNC-53 in fusion with GFP was localised at the microtubule level. Moreover, at lower expression levels, both the centrosomes, and specific plus-end association could be observed. As 10 has been observed with the C.elegans UNC-53 derivatives in fusion with GFP, expressed by the plasmids pEGFPsma and pEGFPecl, the GFP association seems to be less tight as was observed by the full length C.elegans UNC-53 fragment in fusion with GFP. 15 The observed instability of the fusion protein can be due to a lesser association to microtubules, or to a degradation of the fusion protein (figure 43).

20 d) Immunofluorescence on melanoma line G361, and on neuroblastoma line N4 transiently transfected with pLM4.

25 Introduction

25 Northern experiments show that the melanoma cancer line G361 expressed abundantly both the Human1 and Human2 homologue of C.elegans UNC-53. To test if the proteins could be localised in this cell line, a collection of mouse sera was tested on this cell line. 30 To verify if the observation was due to a hu-UNC-53 recognition, and not to an artifact, a positive sera was applied to N4 cells transiently transfected with pLM4, expressing the GFP-hul-Unc fusion. 35

- 71 -

result

a serum, designated 28.1 from a mouse previously injected with peptide (DNRTLPPKKGLYRY) a conserved sequence of the UNC-53 family was used for a immunolocalisation experiment on G361 cells fixed with paraformaldehyde. Antimouse-cy3 was applied as second antibody. Association with microtubule plus-end could clearly be observed. Moreover, in cells showing directional movement, observed as growth cones extensions, abundant staining can be seen in the tip of the growth cone (figure 45). To test whether the recognition of the microtubule associated protein was identical to the Hui-UNC-53 protein, N4 cells were transiently transfected with plasmid pLM4 and consequently fixed with paraformaldehyde and stained with serum 28.1. Only cells that were transfected showed staining with 28.1, indicating that the antibody of 28.1 recognised the Hui-UNC-53-GFP fusion protein (figure 46). This confirms that the staining of the microtubule plus-ends in the growth cones of G361 by serum 28.1 is due to a recognition of at least the Human1 and/or the Human2 homologue. It is concluded that the overexpression of the human homologue of C.elegans UNC-53 in the melanoma cancerline G361 is located on the microtubule plus-ends.

Conclusions

30

- a) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows Unc53 activity
- b) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows microtubule association
- 35 c) - GFP-C.elegans UNC-53 fusion protein

- 72 -

expressed by pEGFP72 shows microtubule plus-end association

5 c) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPec1 show microtubule association.

d) - GFP-C.elegans-UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPec1 no not show microtubule plus-end association

10 e) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPxba and pEGFPsac no not show microtubule associations.

f) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule association.

15 g) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule plus end association.

i) - serum 28.1 recognises the Hu1-UNC-53-GFP fusion protein as expressed by plasmid pLM4 in transiently transfected Neuroblastoma cells N4.

20 j) - the expressed human homologue of C.elegans.- UNC-53 in melanoma line (being at least hu1-Unc-53) is associated with the microtubule plus-ends.

EXPERIMENTAL PROCEDURES

25 Materials

The oligonucleotides used in the PCR-RACE experiments were synthesised by Eurogentee (Belgium). Radioactive compounds were obtained from Amersham.

30 The pCDNA3.1 eukaryotic expression vectors, human 1GT10 cDNA libraries, marathon-RACE cDNAs, human, Northern blots and the T7-tag monoclonal antibody were purchased from Invitrogen. N4, MCF7 and NIH 3T3 cells were retrieved from the Janssen Research cell bank.

35

- 73 -

PCR-RACE conditions

1. A quick screen human cDNA library panel was used to amplify EST clone gb..R41071. The primers used
5 were ESTfw 5'-AATGGCTTCCTGGTTACCTGAG-3' and ESTrv 5'-CAAGTCAGCACCCCGAAGCAGCTCT-3'. Human genomic DNA was used also as template (100ng/reaction). The
10 amplification conditions were as follows: 1 min at 94°C, 30 sec at 55°C, 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72°C. This
PCT fragment was cloned in vector pCR2.1. The resulting plasmid was designed pCR231.

A human heart clone was also produced by RACE-PCR from a human heart Marathon cDNA using the following
15 conditions; 1 min at 94°C, 30 sec at 70°C, 3 min 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72 c KlenTaq DNA Polymerase was purchased from Invitrogen.

For the mouse homologue, total RNA was obtained
20 from N4 murine cells as described. A first strand cDNA was synthesized from 2 µgr of RNA using Ready To-Go cDNA kit (Pharmacia) The primers used were M-ESTfw 5'CCTCTGTGGGCACCGAGGTCACC--3'. The amplification
25 conditions were as follows: 1 min at 94°C, 30 sec at 58°C, 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72°C. All the amplifications product were subcloned in pCRII-1 and several independent clones were analyzed by sequence.

30 2. Screening of Human Heart/Colorectal Adenocarcinoma cDNA library

A human heart cDNA library and a human colorectal adenocarcinoma cDNA library were screened using
35 pCR231bp as probe by the standard plaque hybridization

- 74 -

method. The screening produced several positive clones in each library called respectively λ HH3, λ HH4, λ HH15, λ CAD14 and λ CAD27. The positive phages were purified by two additional rounds of plaque screening and were then amplified.

3. 5' extension using PCR

Three primers with homology to the 5' end of clone λ HH3b were made:

HU53rv1 (5'-cct-ggg-act-gaa-gct-ggt-acc-tga-gcc-3'), HU53rv2 (5'-ttg-gga-aga-gtg-ttc-cga-tcc-cgc-tg-3') and HU53rv3 (5'ggt-gcc-cag-ctc-tgg-ggc-ttc-cac-tcc-3') and used together with λ gt10rv primer (5'-gag-gtg-gct-tat-gag-tat-ttc-ttc-cag-ggt-a-3') in three nested PCR reactions on a cDNA amplified library from Human Heart (Clontech). The reaction mixes contained 25pmol of each primer, 1 mM of each dNTP, 1 μ lKlenTaq Polymerase Mix (50x) and 0.1 ng DNA. The cycling parameters for the first PCR were: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 51°C and 3 min at 72°C and a final extension of 10 min at 72°C, using HU53rv1 and λ gt10rv as primers. 0.4 μ l of this primary PCR product was amplified using HU53rv2 and λ gt10rv as nested primers with the following parameters: 3 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 52°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. The second nested PCR reaction was performed on 0.4 μ l of a 1/50 diluted purified 2.4 kb fragment using HU53rv3 and λ gt10rv as primers: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 56°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. A 774 kb amplification product was subcloned in pCR2.1, resulting in plasmid pCB210-14. The clone fragment was analyzed by sequencing. This fragment

- 75 -

extends 699 bp in 5' direction (see fig 9).

4. 5' extension using PCR

5 Primer HU53rv4 (5'-ccc-tgc-ttg-gtg-ctg-agg-aga-
ctg-g-3') was designed on the 5' end of clone pCB210-
14 and was used together with λ gt10rv to amplify a
fragment of the Human Heart cDNA library with the
following parameters: 3 min at 94°C, 35 cycles of 1
10 min at 94°C, 1 min at 60°C and 3 min 30 sec at 72°C
and a final extension of 10 min at 72°C. A 887 bp
fragment was subcloned in pCR2.1, resulting in plasmid
pCB212. The clone fragment was analyzed by
sequencing. This fragment extends a further 767 bp in
15 5' direction (see fig 9).

5. Human Heart Library screening using the 0.8 kb insert of pCB212 as probe

20 The EcoRI digested and purified clone pCB212 was
used as probe to screen the Human Heart cDNA library
(Clontech) using standard plaque hybridization method.
The positive phages were purified by two additional
rounds of plaque screening. The insert of the λ DNA
25 (produced using Qiagen Lambda Kit) was analyzed by
sequencing. This pHH14-3 resulted in a 2663 bp
fragment overlapping pCB212, pCB210-14 and the 3' end
(434 bp) of λ HH3b and in a 761 bp 5' extension (see
fig 9).

30

3' and 5' extension of HU-Unc53/2 from EST46037

WashU-Merck EST 46037

Transformed cells carrying the EST 46037 sequence
35 were ordered from Research Genetics. Plasmid DNA was

- 76 -

isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

5 3' extension of EST 46037 by RACE

Marathon-Ready cDNAs (Clontech) are premade "libraries" of adaptor-ligated double-stranded cDNA ready for use as templates in RACE experiments.

10 Five ml Marathon-Ready cDNA was used as template in a regular 50ml RACE. The RACE mixture contained 1x KlenTaq PCR buffer, 0.2 mM of each dNTP, 1x advantage KlenTaq polymerase mix (Clontech), 0.15 mM AP1 adaptor primer and 0.15 mM RACE gene specific primer. The
15 amplification conditions were as follows :

94°C for 1 min, 5 cycles of 94°C for 30 s and 72°C for 4 min, 5 cycles of 94°C for 30s and 70°C for 4 min, 25 cycles of 94°C for 30 s and 68°C for 4 min.

20 One-hundred-fold diluted RACE product was used as a template in a nested PCR with AP2 adaptor and gene specific nested PCR primers. Specific nested PCR fragments were cloned into pCR Γ 2.1 (TA cloning kit, Invitrogen) and the sequences of the inserts were
25 determined.

gene specific primer (EST46037-F1)
5'AGTGAGAACAATGCTGTGGACATGC nested gene specific
primer (ES46037-F2) 5'CTGCTCAACTGCAAGTACCACAAATGC
Marathon cDNA library : human placenta

30

WashU-Merck EST 923793

Transformed cells carrying the EST 923793
sequence were ordered from Research Genetics. Plasmid
35 DNA was isolated using standard protocols (Qiagen

- 77 -

plasmid DNA isolation kit), the sequence of the insert was determined.

5 RACE fragments 1.4 and 3.7, 5' extension of
 EST46037

 Method as described previously. Gene specific
 primer (EST46037-R1) 5'ACTGCCTTGAGACTCTGACTTCAGC
 nested gene specific primer (ES46037-R2)
10 5'TGGGCAGAACTGAGAGCTTCTAAGC Marathon cDNA library :
 human placenta

RACE fragments B2.1, D2.1, H2.1; 5' extension

15 Method as described previously:gene specific
 primer (97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG
 nested gene specific primer (97010708)
 5'ACCTGAGTCCTTTCTTAGGCAAAGTGTTCC Marathon cDNA library
 : human placenta (fragment B2.1)
20 human HeLa S3 (fragment D2.1) human colorectal
 adenocarcinoma SW480 (fragment H2.1)

PCR fragments E2.3, C2.3

25 EST 485068 is similar to but not identical with
 the 5'end of HU-Unc53/1. A primer pair consisting of
 one 3' EST 485068 primer and one 5' HU-Unc53/2 primer
 were used to PCR amplify those fragments. lgt10 human
30 placenta Quick screen library (fragment C2.3) or
 Marathon cDNA from human HeLa S3 (fragment E2.3) were
 used as templates in a PCR. A 50 ml reaction mix
 contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM
 MgCl₂, 0.2 mM of each dNTP, 0.15 mM forward and
35 reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer)

- 78 -

and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were purified (Qiagen PCR purification kit) and direct sequenced.

primers :

(97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG

(97012802) 5' CGCTCCCCATCAGATGCAGGCCGG

PCR fragment E1.3-3

EST 01222 is homologous but not identical with the 5'end of HU-Unc53/1. A primer pair consisting of one 3' EST 01222 primer and one 5' HU-Unc53/2 primer were used to PCR amplify this fragments. Marathon cDNA from human HeLa S3 was used as template in a PCR. A 50 ml reaction mix contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM MgCl₂, 0.02 mM of each dNTP, 0.15 mM forward and reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer) and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were analysed on an agarose gel, the fragment of interest was sliced out, purified (Qiagen PCR purification kit) and cloned into pCRIT2.1. The sequence of the insert was determined.

- 79 -

RACE fragments A2.2-2, B2.1-4, D2.1-5; 5'
extension

5 Method as described previously.
 gene specific primer (97041701)

5'TATGCTACGGCCACTCATCTCCGTGG
nested gene specific primer (97041702)

10

5'TGTAACCTGAGTTCCCCTTAAACTGG

Marathon cDNA library :

human placenta (fragment A2.1-2)

human HeLa S3 (fragment B2.1-4)

15 human colorectal adenocarcinoma SW480 (fragment
 D2.1-5)

Translation-initiation splice variants, fragments
D4.1-1, J4.1.4, G4.1.1, F4.1.2

20

Four different translation initiation splice
variants were detected by 5'RACE.

 Method as described previously.
25 gene specific primer (97080803)

5'TCGGTTGTTAGCAGTAGTTGACCCTCC
nested gene specific primer (97080804)

30 5'ACCTGAAAGTCTGGACTGCATTTCAGC
 Marathon cDNA library : human colorectal
 adenocarcinoma SW480 (fragment D4.1-1) gene specific
 primer (97080801)

35 5'ACAACCTGGATAATCTGGGCCAGGAGG

- 80 -

nested gene specific primer (97080802)

5'TCTTGCTGGAGATCCTTGATGAGACGC

Marathon cDNA library :

5

human melanoma G361 (fragment J4.1.4)

human HeLa S3 (fragment G4.1.1)

human placenta (fragment F4.1.2)

10

DNA sequencing

15

PCR amplification products and cDNA clones were subcloned either into pBluescript vectors (Stratagene) or in PCR-IIa vector (Invitrogen) and sequenced either manually by the dideoxynucleotide chain termination method with modified T7 DNA polymerase (Sequenase, United States Biochemical) or automatically with an Applied Biosystems 373 DNA sequencer using the fluorescent terminator kit (Perkin Elmer).

20

RNA blots

25

A Human multiple tissue Northern (MTN-1, Clontech) containing in each lane 2 mg of poly A + RNA from eight different human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas) and a MTN-II human multiple tissue Northern, containing in each lane 2 mg of poly A + RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral leukocyte, were hybridized according to the manufacturer's instructions and washed out in 0.1xSSC:0.2% SDS at 55°C. Also from Clontech, a poly A + RNA blot from human cancer cell lines (melanoma G361, lung carcinoma A549, colorectal adenocarcinoma SW480, Burkitt's

35

- 81 -

lymphoma Raji Leukemia Molt 4, lymphoblastic leukemia K562, HeLa S3 and promyelocytic leukemia HL60) was tested.

5 Construction of plasmids

Plasmid pCDU2 (Figure 17) was constructed by cloning the 2.8 kb *ApaI*-*NarI* fragment from pTB72, the latter restriction site made blunt with klenow enzyme,
10 into pCDNA3, digested with *EcoRV* and *ApaI*. pCDU2 encodes for the homology blocks A, B, C, D and E. Plasmid pCDU3 (Figure 15) was constructed by cloning the 1.9 kb *ApaI*-*NdeI* fragment from pTB72, the latter restriction site made blunt with Klenow enzyme, into
15 pCDNA3, digested with *EcoRV* and *ApaI*, pCDU3 encodes for the homology blocks C, D and E. Plasmid pCDU4 (Figure 16) was constructed by cloning the 1.4 kb *ApaI*-*StyI* fragment from pTB72, the latter restriction site made blunt with Klenow, into pCDNA3 digested with
20 *EcoRV* and *ApaI*. pCDU4 encodes for the homology block E.

Expression of a domain of the human UNC53 in eukaryotic cells

25

1. pCB201: Equivalent construct of human 1 homologue to expression construct pCDU4 of *C. elegans* *unc-53* gene cloned in a eukaryotic His-tag, Xpress Ab tag expression vector.

30

A suitable Bam HI site was engineered on pHH15 open reading frame by amplification with hh15fw primer 5'AGAGCGGATCCATATGCCTCCTTGCCGTCAAGGTG-3' and M13rv primer (5'-cag-gaa-aca-gct-atg-ac-3'). The amplified
35 fragment was then moved to pCDNA3.1.His-A-Vector

- 82 -

digested with BamHI and EcoRI. This new plasmid called pCB201 (Figure 13) produces a cDNA which codes for a fusion protein consisting of a 49 amino acid aminoterminal fragment containing an His-tag and also
5 a T7 epitope tag followed by amino acids 1255 to 1627 of the sequence of the human homologue. pCB201 was also checked by sequence and the n was used in stable transfection experiments carried out in N4, MCF7 and NIH3T3 cells.

10

2. pLM5: Equivalent construct of human 1 homologue to expression construct pCDU3 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector.

15

The phage HH3b was linearized using XhoI. A BamHI and XbaaI restriction site were created on the pHH3b open reading frame using U3-Bfw (5'-cca-cac-tag-ggg-atc-cat-gca-aat-gag-g-3') and U-rv (5'-caa-aag-tct-cta-gag-gag-gcc-agt-3') as primers. This
20 amplified fragment was then moved to pBluescript KS, digested with BamHI and XbaI. Sequencing of this plasmid, named pCB300, showed an amino acid change from a serine to an asparagine due to a change from guanine to adenine on the position 4237 of the DNA
25 sequence. This fault was repaired by cloning a 1418 bp fragment of pLM1 (see below) (using NarI and XbaI as enzymes) into pCB300 digested with the same enzymes. The phage HH3b fragment of this plasmid, named pLM6 (fig 53), was then removed using BamHI and
30 XbaI, to pCDNA3.1/HisA digested with the same enzymes. This new plasmid, named pLM5 (fig 52), produces a cDNA which codes for a fusion protein consisting of a 49 amino acids aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by aminoacid 1069 to
35 1627 of the transcript of HU-Unc53/1. Plasmid pLM5 was

- 83 -

checked by sequencing and used on transient and stable transfection experiments carried out in N4 cells. The plasmid pLM1 was created using a PvuII and partial BamHI digested fragment of pHH14-3 and a BamHI and SpeI digested fragment of phage HH3b, cloned into pBluescript KS digested with SmaI and SpeI. The pLM1 contains the full transcript of HU-UNC-53/1 available at this moment (see fig 9).

10 3.pCB251: Equivalent construct of human 1
homologue to expression construct pCDU2 cloned in an
eukaryotic His-tag, Xpress Ab tag expression vector

15 The phage HH3b was linearized using XhoI. A
BamHI and XbaI restriction site were created on the
pHH3b open reading frame using U2fw (5'-aag-gga-tga-
ttc-ggt-cag-gat-cct-tc-3') and U-rv (5'-caa-aag-tct-
cta-gag-gag-gcc-agt-3') as primers. The amplified
fragment was then moved to pCR2.1. This plasmid was
20 named pCB250. The pHH3b fragment was removed from
pCB250 using BamHI and XbaI and cloned in
pCDNA3.1/HisC digested with the same enzymes. This
plasmid, named pCB251 (figure 55), was checked by
sequencing. pCB251 produces a cDNA which codes for a
25 fusion protein consisting of a 49 amino acid
aminoterminal fragment harboring a His-tag and a T7
epitope tag, followed by amino acids 828 to 1627 of
the partial transcript of HU-Unc53/1. pCB251 was used
on transient and stable transfection experiments
30 carried out in N4 cells (see fig 56).

 4. pLM3: the partial transcript of HU-Unc531
cloned in an eukaryotic His-tag, Xpress Ab tag
expression vector

35

- 84 -

pLM1 was digested with EcoRV and XbaI. This fragment was cloned in pCDNA3.1/HisB, digested with the same enzymes. pLM3 produces a cDNA which codes for a fusion protein consisting of a 49 aminoacid aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by amino acids 1 to 1627 of the transcript of HU-Unc53/1 available at this moment. pLM3 was used on transient and stable transfection experiments carried out in N4 cells.

5. pLM4: the partial transcript of HU-Unc53/1 cloned in an eukaryotic GFP expression vector

pLM1 was digested with ClaI and XbaI. This fragment was cloned in pEGFP-c1, digested with AccI and XbaI. This plasmid was named pLM4. This plasmid produces a cDNA which codes for a fusion protein consisting of GFP, followed by aminoacid 1 to 1627 of the transcript of HU-Unc53/1. pLM4 was used on transient and stable transfection experiments carried out in N4 cells (see figs 43 and 46).

Stable transfection of MCF-7 cells:

Cells were seeded at a density of 2×10^6 cells in a 75 cm² flask using standard culture medium ((Dubecco's MEM, 450 mg/l glucose, 862 mg/l L-Alanyl-L-Glutamin, 110 mg/l Na-pyruvate; GibcoBRL) supplemented with 10% foetal calf serum (FCS; GibcoBRL), and 100 U/ml penicillin (GibcoBRL) and 100 µg/ml streptomycin). The culture was grown at 37°C in a 10% CO₂ atmosphere, to approximately 70% confluency (approximately 18 hours). The culture medium was removed and 10 ml MEM-HEPES (GibcoBRL) supplemented

- 85 -

with 10% FCS was added to the cells. The culture was further incubated for four hours at 37°C in standard sterile air. DNA-CaCl₂ was meanwhile prepared by mixing 30 µg DNA in 0.1 x TE (1 mM Tris. Hcl, pH 7.2, 0.1 mM EDTA, pH 8) and 0.1 ml 1.25 M CaCl₂/HEPES (1.25 M CaCl₂, 0.125 M HEPES; pH 7.05). 0.1 x TE was added to a final volume of 0.5 ml. The DNA-CaCl₂ was added drop by drop to 0.5 ml BS/HEPES (25 mM HEPES, 0.25 M NaCl, 0.01 M KCl, 1.4 mM Na₂HPO₄, 0.01 M glucose, pH 7.05) while pipeting a sterile airflow through the latter solutions. The DNA-Ca₃(PO₄)₂ precipitate was then placed at 37°C for ten minutes. The DNA-Ca₃(PO₄)₂ precipitate was vortexed and added to the cells, together with 100 µl of a 0.01 M chloroquine (Sigma) stock in H₂O. After four hours of incubation at 37°C in sterile standard air, the medium was removed, and the cells were washed with PBS (GibcoBRL). 25 ml of medium was added and the cells were placed at 37°C in a 10% CO₂ atmosphere. After 48 hours of incubation, the cells were harvested, diluted and cultivated under selection (600 µg/ml G418 (Duchefa)) for two weeks prior to clone selection. Mock transfected MCF-7 were positive for the beta-galactosidase transgene. The stability of transfection in MCF-7 was assessed by passaging cells four times in the absence of Geneticin and then re-exposing them to the selector agent. In these experiments, unc-53 or mock transfected cells proliferated, whereas untransfected MCF-7 cells proliferated at a much slower rate.

Stable transfection of N4 neuroblastoma cells

Cells were seeded at a density of 2x10⁶ cells in a 25 cm² flask using standard culture medium ((MEM Rega 3; GibcoBRL) supplemented with 10% FCS, 0.14%

- 86 -

Na₂CO₃,
2 mM glutamine, 100 U/ml penicillin, and 100 µg streptomycin). The culture was grown overnight at 37°C in a 10% CO₂ atmosphere. Transfection mixture was prepared by adding 12 µg DNA in 600 µl opti-
5 1(GibcoBRL) to 36 µl LipofectAMINE (GibcoBRL) in 600 µl opti-
10 1. This was done by adding drop by drop the first solution to the second. The mixture was placed for 30 minutes at room temperature, after which 1.8 ml of opti-
15 1 was added. In the meanwhile the cell culture was washed twice with opti-
3 ml of transfection mixture was added. The culture was placed at 37°C in sterile standard air. After four hours, 3 ml of normal culture medium was added and the culture was placed at 37°C under 10% of CO₂.
18 hours later, the culture was washed with PBS, and fresh normal culture medium was added. A further 24 hours later, the cells were harvested, diluted and cultured under selection (750 µg/ml G418) for two
20 weeks prior to clone selection.

Fixation of cells for Immunofluorescence

Medium was removed from the 9 cm² wells containing the coverslips. A 4% solution of
25 paraformaldehyde (Sigma) in PHEM (1 g/l glucose, 0.4 g/l KCl, 8 g/l NaCl, 0.06 g/l KH₂PO₄, 0.0475 g/l Na₂HPO₄, 0.35 g/l NaHCO₃, 1.51 g/l PIPES, 0.76 g/l EGTA, 0.19 g/l MgCl₂; pH 6) was added for 30 min at
30 room temperature. The fixative was removed, and the coverslips were washed three times 10 minutes with PHEM. The coverslips were then placed in PHEM, containing 0.5% Triton-X100 (Serva) for 30 minutes, after which the slide was washed again for three times
35 10 minutes with PHEM. The coverslips were then placed

- 87 -

under PBS (0.14 M NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.3) containing 0.2% Tween (Sigma) for at least one hour at 4°C

5 Immunofluorescence staining

10 The coverslips were inverted on 35 µl of appropriately diluted antibody, being YL 1/2 for tubulin and/or mab 16-48-2 monoclonal or anti-UNC53 (gp48) polyclonal antibody for UNC53. The slides were placed at 4°C for at least 18 hours. Excess of primary antibody was then removed by washes of three times ten minutes in PBS-Tween. The slides were then treated with secondary antibody in the same way as for 15 the primary antibody. F-actin was labelled by including TRITC- or FITC coupled phalloidine to the incubation buffer. The inverted slides on the secondary antibody were left at room temperature for approximately one hour. Slides were then washed again 20 for three times ten minutes with PBS-Tween and once with PBS. The coverslips were mounted on slides with the medium described by Herzog et al. (Cell Biology: a laboratory handbook, 1994, Academic Press, 355-360). After at least two hours, slides were ready for 25 analysis.

Time lapse analysis

30 Analysis of the behaviour and movement of growing cell cultures was done by placing a non-confluent culture under a phase contrast microscope equipped with a temperature controlled stage (37°C). Images were recorded using a CCD camera (COHU 4912) coupled to a SCION LG3 framegrabber in a Macintosh ppc 8100 35 running NIH image version 1.60. Images were recorded

- 88 -

at time intervals, varying from 15 sec to 1 min. for half an hour to two hours. Image enhancement and playback was done in NIH image.

5 Phagokinesis

A variety of cell types were shown to migrate over colloidal gold coated culture plastic or glass and displace or phagocytose the gold lawn on their way while locomoting. The track left bare is a qualitative and quantitative measure of cell motility and/or locomotion. The basic methods have been described in detail elsewhere (Albrecht-Buehler, 1977, Cell, 11: 395, Zetter, 1980; Nature, 285: 41; O'Keefe et al., 1983; J. Invest. Dermatol., 85: 130). Culture plates were gelatin and gold coated as described by Albrecht-Buehler (1977). Unc-53 and mock transfected MCF-7 were seeded in plates at low density and allowed to adhere to the plate and to locomote overnight. Cells were chemically fixed to the plate, washed and air-dried. Images of the gold lawns were captured using automated videomicroscopy; composite images of the wells were generated and single-cell phagokinesis tracks were measured using a home-made routine in SCIL™ software.

C. elegans-UNC-53 preferentially binds microtubule plus ends or GTP-tubulin

30 1. Cloning of C.elegans cDNA in pEGFP-C1 and construction of C-terminal deletion derivatives.

a) Constructing a GFP-Unc53 N-terminal fusion:

35 A PCR experiment was performed under standard conditions, using pTB72 as template and cp17

- 89 -

(ata gcc aga tct acg tca aat gta gaa ttg) and cp18 (ttt aga aac cgc ggg tgg) as primers. The resulting 0.4 kb fragment, coding for the N-terminal fragment of C.elegans Unc 53 was cloned in vector pCR2.1 (original TA cloning kit, Invitrogen), resulting in plasmid pTA1718. The 0.4 kb fragment was isolated as a BglII-SacII fragment and cloned in pEGFP-C1 (Clontech) digested with the same enzymes. The resulting plasmid was designated pEGFPsac (Figure 29). pEGFPsac encodes the N-terminal 13 aa of C.e.Unc53 in fusion with GFP.

b) Construction of a GFP-C.e. Unc53 full length fusion:

Plasmid pTB72 (shown in Figure 1) was digested with restriction enzymes SacII and ApaI. The resulting 4.5 kb cDNA fragment, encoding for the C-terminal fragment of C.elegans Unc53 was cloned in plasmid pEGFPsac (Figure 29), digested with the same enzymes, resulting in plasmid pEFP72 (Figure 30). Plasmid pEGFP encodes GFP in fusion with the full length C.e. Unc53.

c) Construction of N-terminal deletions of GFP-C.elegans UNC-53 fusion protein, other than pEGFPsac:

pEGFP72 was digested with SmaI. The resulting 7.0 kb fragment was religated and transformed in E.coli, resulting in plasmid pEGFPsma (Figure 31). This plasmid codes for the first 760 aa of the Ce-UNC-53 in fusion GFP.

pEGFP72 was digested with restriction enzymes Ec1136II and SmaI, the resulting plasmid after ligation and transformation in E.coli of the 6.7 kb fragment was designated pEGFPec1 (Figure 32). This plasmid codes for the N-terminal 670 aa of the C.e. Unc53 in fusion with GFP. pEGFP72 was further digested with SmaI and XbaI. The latter site was made blunt with Klenow polymerase. The resulting fragment

- 90 -

of 5.4 kb was religated and transformed in E.coli. The resulting plasmid was designated pEGFxba (Figure 33). This plasmid codes for the N-terminal 256 aa of C.elegans Unc53 in fusion with GFP.

5

2. Constructing a hu1-UNC-53-GFP fusion, and a deletion derivative

10 The 5.4 kb hu1-unc53 fragment was isolated as ClaI-XbaI fragment from pLM1 (Figure 54), and cloned in pEGFP-C1 digested with AccI and XbaI. pEGFP-C1 was isolated from E.coli GM41 (Hfa H, dam-3, thi-1, rel-1). This makes the XbaI restriction site available for restriction digest. The resulting plasmid was
15 designated pLM4 (Figure 34).

3. Visualisation of GFP fluorescence in N4 cells

20 N4 neuroblastoma lines where seeded in Lab Tek chambered coverglass (Nalge Nunc International) and transfected using lipofectAMINE (GibcoBRL). After 18 hours, the chambered coverglasses where placed on a inverted microscope, and GFP fluorescence could be
25 visulalised.

4. Staining GFP fusion expressing cells with antibodies

30 Transfection with the GFP fusion constructed was also performed on coverglasses in a 6-well plate. After paraformaldehyde or methanol-acetone fixation, cells could be stained for actin cytoskeleton with TRITC-phalloidine, for hu-unc53 with sera 28.1 and for
35 tubuline with YL1/2 antibody. Visualisation was then

- 91 -

performed on a axioplan (Zeiss microscope).

5 Methods of Producing and Observing the Effects of
A Chimeric unc-53 Gene

1. Definition of a promoter region in the unc-53 C.elegans gene:

10 The genomic region from the position 15621 to
18415 in the C.elegans unc-53 gene, called promoter A,
was cloned and fused to the cDNA of the GFP gene
(clone pA/GFP, or pNP10)(cf. fig.51). This construct
15 is injected into wild type worms (N2). Transgenic
line express GFP in different neurones: the two pairs
of pioneering neurones PVP and PVQ, both BDU neurones,
both ALN and PLN neurones, both PDE neurones, both PVM
neurones, and 4 vulval cells. Expression begins in
early embryogenesis, when the axons of those neurones
20 grow out.

2. Mutant Phenotype in Unc-53(n152) alleles:

25 In wild type worms (N2), the two pairs of ALN and
PLN neurones each send an axon in a straight line
longitudinally from the tail to the head (see
fig.50a). In unc-53(n152) alleles, the axons are
shorter and often branch in a dorso ventral direction
30 (see fig.50b). The neurones are visualised with the
construct pA/GFP, injected in unc-53(n152) worms.

3. The minigene pA/unc-53 rescues the
elongation defect of ALN and PLN neurones:

35

- 92 -

The promoter A from the *C.elegans* unc-53 gene was fused to the cDNA of the *C.elegans* unc-53 gene (clone pA/unc53, or pNP9). This construct was injected in unc-53(n152) mutant worms, together with the pA/GFP construct described above to visualise the ALN and PLN neurones. The elongation defects of those neurones in the unc-53 mutant are almost completely restored by the expression of the unc-53 cDNA express under the promoter A (see figs. 50 and 51b).

4. Domain swap between the *C.elegans* and human unc-53 gene:

To test whether the vertebrate and worm members of the unc-53 family are functionally equivalent, we tested the ability of the human gene to rescue the mutant phenotype in the worm. We replaced the carboxyterminal predicted nucleotide binding domain (NTPase) of the worm protein with the homologous fragment of the human 1 gene.

The clone pA/unc-53 was deleted of the *C.elegans* NTPase domain, from the HpaI site, position 29800 on the genomic of unc-53, and replaced by the equivalent domain of the human-1 gene (unc-53H1) (see fig. 51). The resulting clone is named pA/unc-53H1. When this clone is injected to unc-53(n152) mutants, the transgenic worms show a significant but incomplete rescue of the defect in the elongation of the ALN and PLN neurones (see fig. 51b). The axons are longer, often elongated until the region of the vulva in a straight line, without branching dorsally anymore. This result shows that a NTPase region of the human unc-53 homologue can functionally replace the NTPase region of the *C.elegans* worm.

- 93 -

The degree of rescue was analyzed quantitatively and summarized in Figure 51b:

The four strains compared are:

wt; un-53(n152); unc-53(n152), pA/unc-53; unc-
5 53(152), pA/unc-53-H1.

The various phenotypes observed are brought together in three large classes:

<<wild type>> the axon is straight, unbranched and migrates into the head;

10 <<vulva>> the axon is straight, unbranched and stops in the vulva region;

<<mutant>> the axon is short, does not reach the vulva region and has collateral branches.

The figures are indicated as a percentage. The number
15 of axons observed is indicated in the following column.

The data clearly show demonstrate conclude that the nematode/human chimera minigene pA/unc-53-H1 partly rescues the defects of the axonal migration of
20 the ALN and PLN neurones and demonstrate conservation of function of this domain between man and worm. The transgenic lines provide a functional screening assay for the motility function of at least part of the human UNC-53 gene.

25

II. Materials and methods

1 - Cloning:

30 a) pAB/GFP (pNP3 - Figure 27)

The gene of GFP has been amplified by PCR with cpn3 oligo-nucleotides

"acattaagcttcgtacgcttgagggtaccg" and Cpn5
35 "gaaaggatccgtacgataaggtattttgtgtcgg" on the plasmid pPD95.75(Figure 59) so as to be inserted at the 5'

- 94 -

position in fusion into the exon 12 of the unc-53 gene at a single restriction site SphI and contains its stop codon at 3' plus one polyadenylation site. The PCR amplification product is directed by HindIII and BamHI, sites which are contained respectively in the cpn3 and cpn5 oligonucleotides and sub-cloned in the pBS vector (clone pNP2). The GFP is then excised from the pNP2 clone at the site SphI and integrated into the X16 clone (Figure 60) originating from sub-cloning of the lambda phage S4 digested by XhoI. The X16 clone containing the genomic sequence of unc-53 from the position 16621 to the position 24891 cloned in the site XhoI of pBS.

b) pAB/unc-53 (pNP8 - Figure 35)

The promoter region AB of the X16 clone (between PstI and SphI) has been inserted in the clone pTB115 (Figure 58) in which the region between the sites PstI and SphI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

c) pA/GFP (pNP10 - Figure 56)

The promoter region A has come from the X16 clone between the sites PstI and NheI and integrated in the vector pPD95.75 containing the GFP in the sites PstI and XbaI.

d) pA/unc-53 (pNP9 - Figure 44)

The promoter region A has come from the X16 clone between the sites PstI and BstXI and is integrated into the clone pTB115 in which the region between the sites PstI and BstXI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

e) pA/unc-53 -H1 (pCB501 - Figure 57)

- 95 -

The clone pA/unc-53 (pNP9) has been deleted from the region 3' of the gene unc-53 of the nematode between the sites HpaI and NcoI. The 3' region of the Hlunc-53 gene has been amplified by PCR with the oligonucleotides U4Afw (5'-gca-cat-cgt-taa-cgg-gga-ctt-gaa-gc-3') and Urv (5'-caa-aag-tct-cta-gag-gcc-agt-3') and digested with HpaI and XbaI. After a filling stage with T4 polymerase, the ligation is effected with a complete end.

2-Injection

Conventional injection techniques are used (Fire A., 1986, Mello G, et al, 1991, journal Mello G. and Fire A., 1995). Young hermaphrodite adults are injected in their two syncytial gonads. The DNA used is prepared in standard manner (Qiagen) followed by precipitation with lithium chloride. After an extensive rinsing stage to eliminate all the salts, the DNA is resuspended in water. The injection solution contains the different DNAs at a concentration of 100 ng/μl in an injection buffer:. The plasmid pRF4 containing the dominant allele su 1006 of the gene rol-6 (Kramer J. et al, 1990, Mello C. et al, 1991) is used as a transformation co-marker. The descendants of roller phenotype of the hermaphrodite injected are isolated. Approximately 10 % of these transformants will yield a stable strain, in which the different DNAs injected are associated to form a mini-chromosome which will segregate as unstable extrachromosomal arrays. All the transgenic strains obtained were tested by PCR for the presence of the DNA injected, using a specific primer of the vector and a primer in the gene (results not shown).

3. Microscopy

- 96 -

The nematodes are observed under a ZEISS Axioplan microscope provided with Nomarski lenses, with 40X Neofluar, 63X Plan-Apochromat, 100X Plan-Apochromat objective lenses. For fluorescence observation the luminous source is a mercury bulb. Different ZEISS filters are used:

- for observation under GFP fluorescence, FITC filter: blue excitation line at 588 nm, emission through a 515-565 nm band-pass filter;

- for observation of the antibody labelling with a secondary antibody coupled to the TRITC: excitation through a 546 nm band-pass filter, emission through a 590 nm long-pass filter.

The image acquisition is effected by means of a CCD camera and and NIH image program using a Machintosh computer. The images are processed using the Adobe Photoshop program.

20

25

30

35

Sequence Listing

The following sequences are referred to in the specification:

5

Sequence ID No 1 is an amino acid sequence of human homologue 1 of UNC-53 protein illustrated in Figure 9b.

10

Sequence ID No 2 is an amino acid sequence of human homologue 2 of UNC-53 protein illustrated in figure 11d.

Sequence ID No 3 is a nucleic acid sequence of the hu-1-unc-53 gene illustrated in Figure 9b.

15

Sequence ID No 4 is a nucleic acid sequence of the hu-2-unc-53 gene illustrated in Figure 11d.

Sequence ID No 5 is a nucleotide sequence of Phage Lamda Clone 3b deposited under Accession No LMBP 3595.illustrated in Figure 9.

20

Sequence ID No 6 is a nucleotide sequence of plasmid pLM1 deposited under Accession No LMBP 3762 and illustrated in fig 54.

Sequence ID No 7 is a nucleotide sequence of plasmid pLM4 deposited under Accession No 3763 and illustrated in fig 34.

25

Sequence ID No 8 is a nucleotide sequence of plasmid pEGFP72 deposited under LMBP Accession No 3764 and illustrated in fig 30.

30

Sequence ID No 9 is a nucleotide sequence of plasmid pCB501 deposited under Accession No 3765 and illustrated in fig 57.

Sequence ID No 10 is a nucleotide sequence of plasmid pCB201 deposited under Accession No. LMBP 3594.

SEQ ID No. 1

F. amino ACID SEQUENCE OF
HU-UNC-53/1 PROTEIN

Page 1

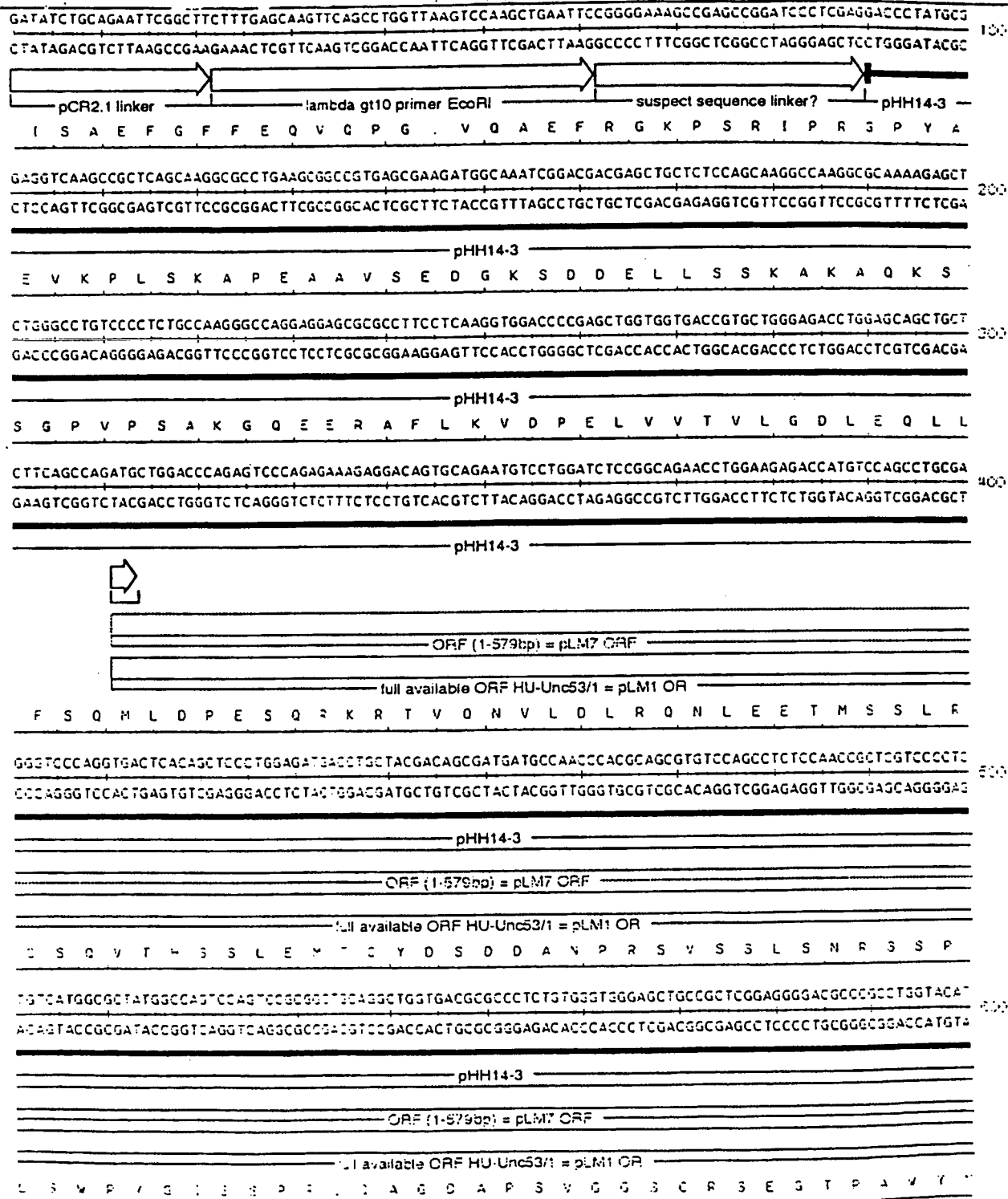
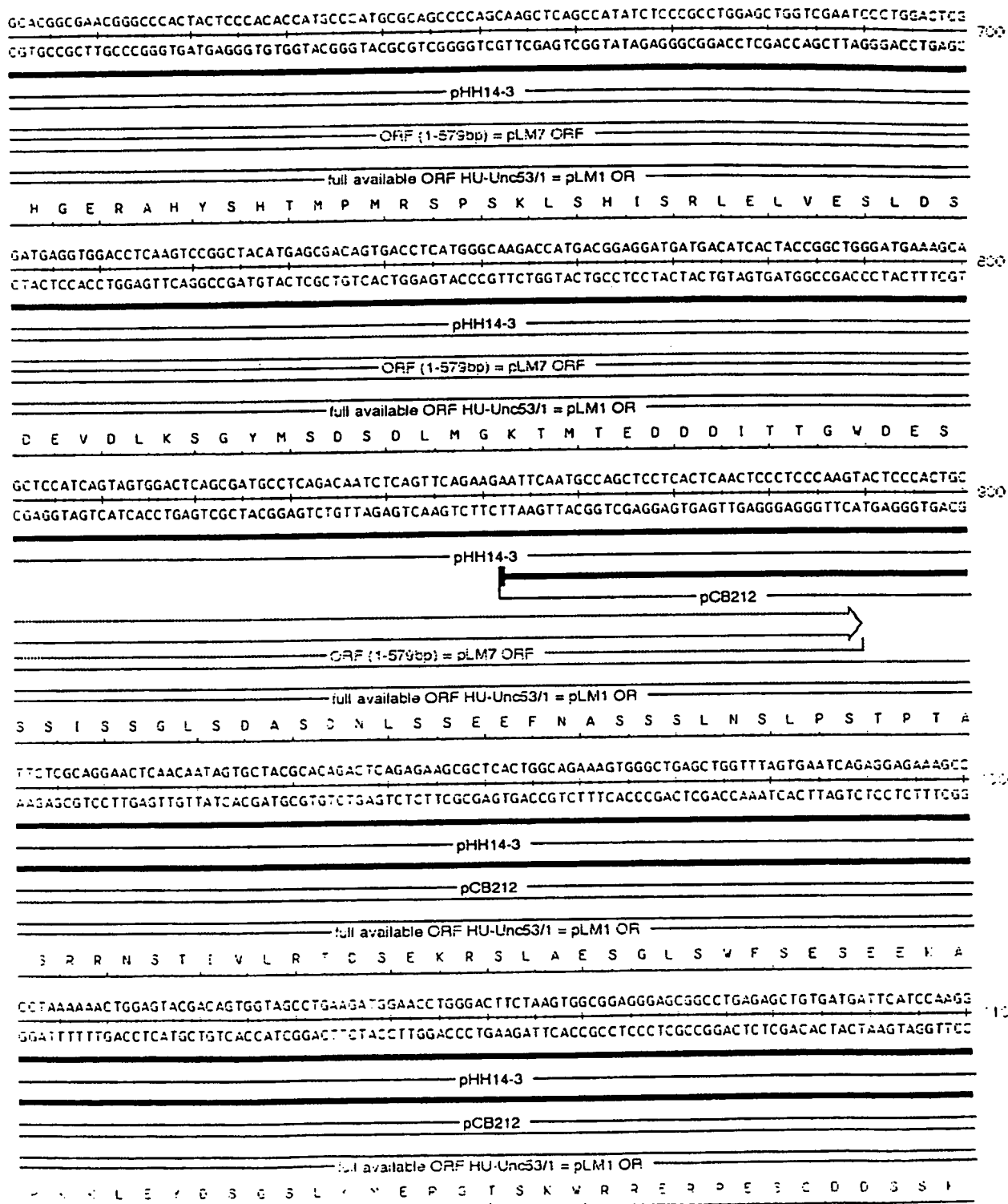


Fig 9

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 2



Tuesday, 18 November 1997 10:33
lig Hu-Unc53/1 seq (1 > 8013) Site and Sequence

Fig 9

Page 3

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCCTGGTTCCCTGAAGAAGGGCAAGACCCACCCTGGGCTGTAACTTCCCCCATCACTCACACAGC
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCTG

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCTCAAAGTCGCAGGCAAACCTGAGGGCAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT
GCTCTCACGGGAGTTTACAGCTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACCTCTTATGACCCGAGGTTGCGAGGAGGAGACTA

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGGAGCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTGATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCGCTGTCGGTGACAGTACGTTTGACCACCAAGTCSTGAGAGTCGTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCCGGTCTCT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R F T

TACCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCTTGGCTCTGGAGCCCGTCTAACAATCCAGTACCGCAGCCTGCCCGGCCAGCCAAGTCAAGT
ATCGAATCTACAAAGTTGTCAGTCTCGGTCCTAAGSACCGAGGACCTCGGGCAAGATTGTAGGTTCATGCGCTCGGACGGGGCCGGTTCAGTTCAT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

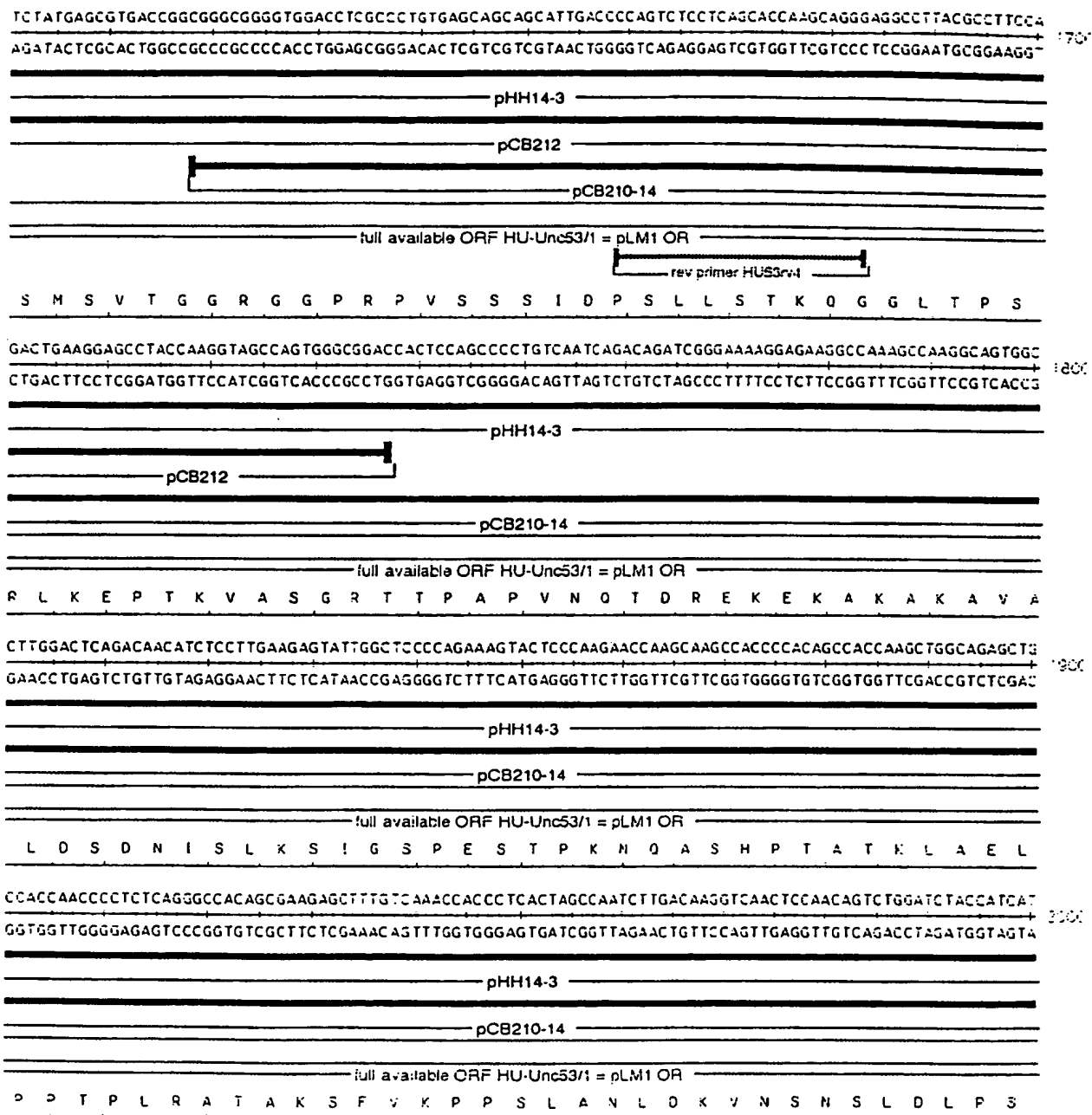
S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A S S

Fig 9

Tuesday, 18 November 1997 10:33

Page 4

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

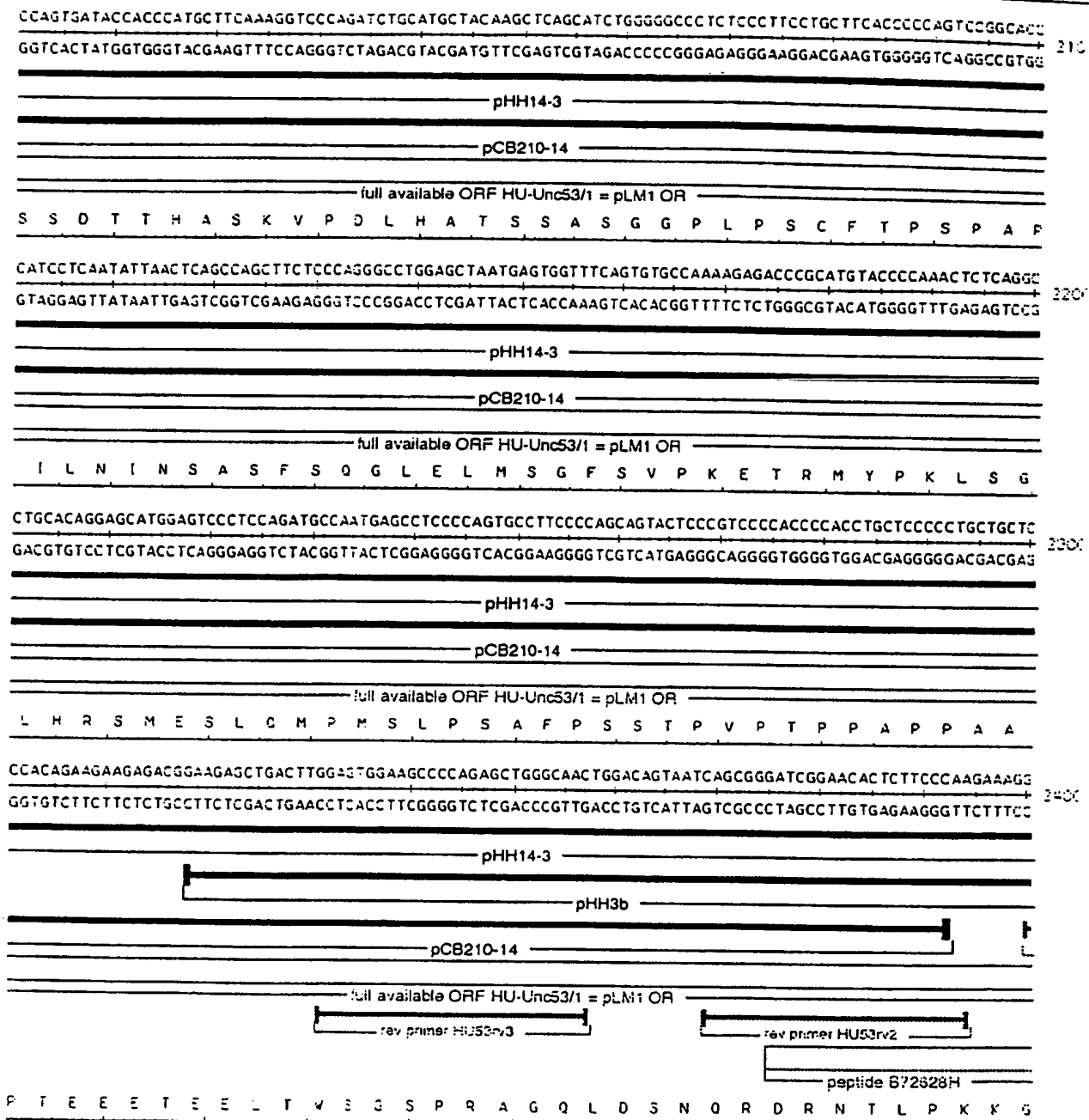


Tuesday, 18 November 1997 10:33

fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 5



Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 6

GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT 250
CGAGTCCATGGTCTGAAGTCAGGGTCTCTCTGGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGC TAC TGGTCAGTC TCGACGGG

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACCTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCAAGCAAGAATCACCCGCTCCAACT 260
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCGGTGGTGCGGTTCTTAGTGGGCGAGGTTGT

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S H

GCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAGCGGCTCCCAATGGGGAGCACCTGTCCC TGGCCGAGAGACCCAAGGGAATGATTCCGTCAGG 270
CGTAGGGGTGGGTGCTCCGCGGAAGCTCGACATGTCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G

ATCCTTCGAGACCCACGGACGATGTTACGGGTCAGTCTGTCCCTGGGCTCCAGTGCCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCT 280
TAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCAAGACAGGGACCGAGGTCACGGAGGAGGTGATGAGGAGTCGACTCCTCTCCTACGT TAGA

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T D C V H G S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAAAATCCGGAAGCTTCGTAGGGAAC TGGAAATCATCCAGSAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT 290
CTCGTTTAGGCTTCGAAGCATCCCTTGACCTTAGTAAGGTCCTTTTACCAGGTTGGAAC TGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I R < L R R E L E S S Q E K V A T L T S C L S A N A N L V A A

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 7

Fig 9

TTGAGCAGAGCCTGGTGAAATATGACATCCCGCCTGCACACCTGGCAGAGACGGCCGAGGAGAAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGA
AACTCGTCTCGGACCACCTATACGTAGGGCGGACGCTGTGGACCGTCTGCCGGCTCCTTCTCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT

300

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAC
GAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTCATAGTCCCTCGGAATTACGGAGTCTTTGGTGTGGGTTCTTGAAGCCTAGTCTCTGTTTTG

310

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I X R Q H

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTCTTTTCTTTTCTCAACCC

320

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S V

TCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCATATAAAAAAGGGGCCAAGTCAGCTTCCTCATACCGGATATAGAGGAGATTGCTACACCGGA
AGATACTCGAAGCTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATGTGSGGT

330

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P G

CTCTCAGCCCCCTCATCCCCAAACTACAGCATGGTCTACAGAGACTGCTTCACCCCTCCATCAAGTCTCCACCTTGTCTCGGTGGGCACTGATGTC
GAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGATGTCCTGACGAAGTGGGAGGTAGTTTCAGGAGGTGGAAACAGGAGGCACCCGTGACTACAG

340

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q H S S T E T A S P S I K S S T L S S V S T D V

Fig 9

Tuesday, 18 November 1997 10:33

Page 8

fir HU-Unc53/1 seq (1 > 6013) Site and Sequence

ACCG .3CCCTGCTCACCCAGCCCCACACTAGGCTGTTCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGGCTCTGAGG
TGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCCTCTCGGTCTCTTCTTCCCTCCATAGCCTCGACGCGAGACTCG 3500

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGT
ATACCCTCTTCTTTACTTTCGAATGCTGTAGGCGAACCCTCGGGAGTGTAGACGGGTGGTTGACCTAGTTCGAAGCCCTCTGGTACGTGTTGTACGTCAA 3600

U2 ORF = pCB251 ORF

pHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCAC TCCAGGGCAGGTCCCTGGATCATCTGCATTA
CCTCCACCTGGACGACTTTCTGCTCTTACTGGCTGACTTCCATCGGGG TCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAA 3700

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCCACGCCCTCCCTAGGCTTGGCACTCACCCATTCCTTCGGCCCCAGTCTTGAGACACAGACCTGTACCCATGGATGGCATCAGTACTTGT
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGSTAAGGAAGCCGGGTCAGAAGCTCTGTGTC TGGACAGTGGGTACCTACCGTAGTCATGAACAG 3800

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T S F G P S L A D T D L S P M D G I S T C

Tuesday, 18 November 1997 10:34

fig. 'u-Unc53/1 seq (1 > 6013) Site and Sequence

Page 9

GTCCAAAGGAGGAAGTGACCC TCCGGGTGGTGGTGAAGTGC CCCCCGAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCTCTGGGCTGTAG
CAGGTTTCTCTCTTCACTGGGAGGCCACCACCAC TCCTACGGGGGCGTCGTGTAGTAGTTTCCCCGTAACCTCGTCGTCCTTAAGAAGGACCCGACATC

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCGAAGTGTCAAGGACTATATTTCTAAATGGACCCAGCCTCTACCCTGGGA
GTTCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGAAGTTCCGTATATAAAGATTTTACC TGGGTGGGAGATGGGACCT

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCGAGATGCCTCCTTGCCGTGAGGTGTCAATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGCTGCACTTGTCTACAACCTACGTCCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

ACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAA TCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCT
TGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGACGCTGCGGACCAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGATTTGGGA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

H I S V S L K G L K E C V D S L V F E T L P K P H M Q H I S L

Tuesday, 18 November 1997 10:34
fig. hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 18

CCCTGGAAGCACCGCGCCTCGTCCCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACC TGSTGGAGCGCTCTGGG
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTCGCCGTGCCCGTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACC
GCATCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAAGCTTCCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGS

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E V T E S I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAACAGGAATTGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGC TCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA
GGTTTGTCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCGAGGTAGTCACTCAACCACTTACCCCGGAGTGGAGCTT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E T G I G D V P L V L L D D L S E A G S I S E L V N G A L T C I

Tuesday, 18 November 1997 10:34
fig. 4u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 11

G T A T C A . A A A T G T C C C T A T A T T A T A G G T A C C A C C A A T C A G C C T G T A A A A A T G A C A C C C A A C C A T G G C T T G C A C T T G A G C T T C A G S A T G T T G A C C T T C C C
C A T A G T A T T T A C A G G S A T A T A A T A T C C A T G G T G G T T A G T C G G A C A T T T T A C T G T G G G T T G G T A C C G A A C G T G A A C T C G A A G T C C T A C A A C T G G A A G A G G

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72525H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

A A C A A C G T G G A G C C A G C C A A T G G C T T C C T G G T T C G T T A C C T G A G G A G G A A G C T G G T A G A G T C A G A C A G C G A C A T C A A T G C C A A C A A G G A A G A G C T G C T T C
T T G T T G C A C C T C G G T C G G T T A C C G A A G G A C C A A G C A A T G G A C T C C T C C T T C G A C A T C T A G T C T G T C G C T G T A G T T A C G G T T G T T C C T T C T C G A C G A A G

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

G G T G C T C G A C T G G G T A C C C A A G C T G T G G T A T C A T C C A C A C C T T C C T T G A G A A G C A G A C C T C A S A C T T C C T C A T C G G C C C T T G C T T C T T C T G T C
C C C A C S A G C T G A C C C A T G G G T T C G A C A C C A T A G T A G A G G T G T G G A A G G A A C T C T T C G T G T C G T G G A G T C T G A A G G A G T A G C C G G G A A C G A A G A A G A C A G

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L D V V P K L W Y - L H T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34
fig 1. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 1

Page 12

GTGTCCCATTTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACCTGTGGAACAACCTCTATCATTCCTTATCTACAGGAAGGAGCCAGGGATGGGATAAAG
CACAGGGTAACCGTAACCTCTGAAGGCCGGACCAAGTAACCTGGACACCTTGTGTGAGATAGTAAGGGATAGATGTCTCTCTCGTTTCTACCTATTTCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCAAC
CAGGTACCTGTCTTTTCGACGAACCTCTCTGGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTGTTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V F G C K A A V E D P V E W V R D T L P W P S A Q Q D C S X L Y H

TCCCCCAACCCACCGTGGGCCCTCACAGCATTGCTTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGA
ACGGGGGTGGGTGGCACCCGGGAGTGTCTAACGGAGTGGAGGGCTCTATCTGTGCAAGTTCTGTCTGGTGGTTCAGAGACCTGAATCTAGSAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L P P P T V G P H S I A S P P E D R T V K O S T P S S L D S C P L Y

Seq ID No 2: Amino Acid Sequence of Hn-wc-53/2 Protein

[illegible][illegible]

FD-302 (Rev. 11-27-70)

[illegible]

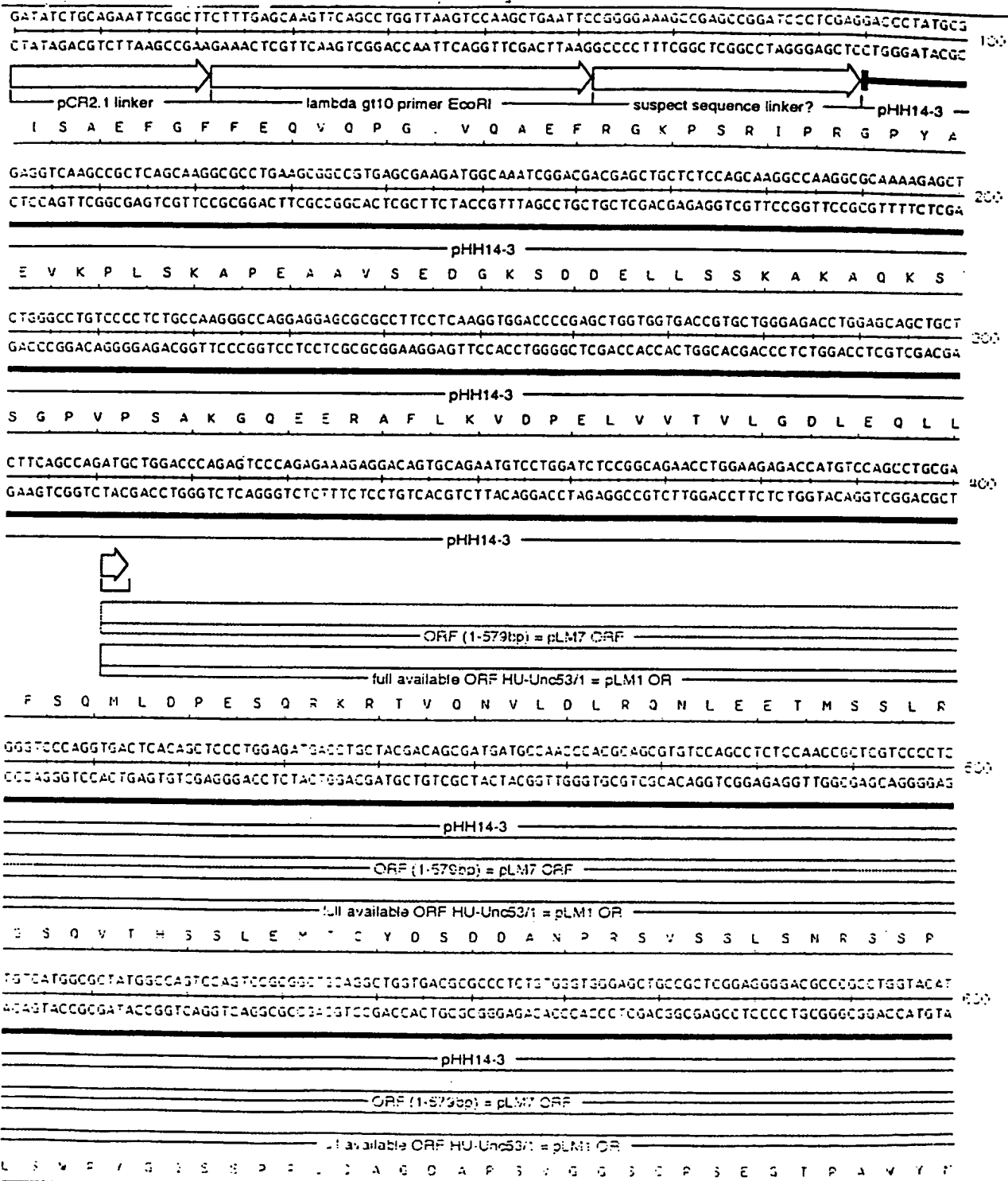
10-11-19

[illegible]

SEQ. ID No 3

Nucleic Acid Sequence of
hu-unc-53/1 gene

Page 1



Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and sequence

Fig 9

Page 2

GCACGGCGAACGGGCCCCACTACTCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCGAATCCCTGGAAGTGG 700
CGTGGCGCTTGGCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCTGTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGG

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L D S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCA 800
CTACTCCACCTGGAGTTCAGCGCGATGTACTCGCTGTCACTGGAGTACCGGTCTCGGTACTGCCCTCTACTACTGTAGTGATGGCCGACCTTACTTTGGT

pHH14-3

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D E S

GCTCCATCAGTAGTGGAAGTCAAGCGATGCCTCAGACAATCTCAGTTCAAGAAGATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 900
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTAGAGTCAAGTCTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGAGC

pHH14-3

pCB212

ORF (1-579bp) = pLM7 ORF

full available ORF HU-Unc53/1 = pLM1 OR

S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P T A

TTCTCGCAGGAAGTCAACAATAGTGGTACGGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAAACAGAGGAGAAAGCC 1000
AAGAGCGTCCCTGAGTTGTATCAGGATGGGTGCTGAGTCTCTCGCGAGTGACCGTCTTACCCCGACTCGACCAATCACTTAGTCTCCTCTTTGGG

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

I R R N S T I V L R T C S E K R S L A E S G L S V F S E S E E I A

CCAAAAAAGTGGAGTACGACASTGGTACGCTGAAATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCATCCAAAGG 1100
GGATTTTGTGACCTCATGCTGACCATGGGAGTTCTACCTTGGACCTTGAAGATTACCGGCTCCCTCGCGGAGCTCTGACACTACTAAGTAGGTTCC

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

P L E Y C S C S L N E P G T S N W R R E R P E I C D D S S I

Tuesday, 18 November 1997 10:33

Fig 9

Page 3

fig HU-Unc53/1 seq (1 > 6013) Site at sequence

GTGGAGAAC TGA AAAAGCCCATCAGCCTGGGCCACCTGGTTCCCTGAAGAAGGGCAAGACCCACCTG TGGCTGTAACTTCCCCCATCATCACACAGC
CAGCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCTCAAAGTCGCAGGCAAACTGAGGGCAAGCTACAGACAAGGGTAAGCTTGCAAGTGAAGAATACTGGGCTCCAACGCTCTCTCTGAT
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCTTCGATGTCTGTTCCTTCGAACTGCAACGTCACCTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCAATTGCTCGCCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCTGCGCA
CGACCAGCCCTGGCGGACTCACTACGATTCTTCSGGGGGAGCCGTAAAGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTGTCATSCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCCGTGTCGGTGACAGTACGTTTACCACCAAGTCCGTGAGAGTCGTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCAATTACCCGCGTCTCT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M C T G G S A T L S K I O K S S G I P V K P V N S R T

TAGCTTAGATGTTTCCAAAGTTCAGAGCCAGGATTCCTTGGCTCCTGGAGCCCGTTCTAACA TCCAGTACCGCAGCCTGCCCGGGCCAGCCAAAGTCAAGT
ATCGAATCTACAAAGGTTGTACAGTCTCGGTCTTAAAGACCGAGGAGCTCGGGCAAGATTGTAGGTTCATGGCGTCGGACGGGGCCCGTGGGTTCAGTTC

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

L L U V S N S A E P I F L A F G A R S N I O Y R S L P R P A S

Fig 9

Tuesday, 18 November 1997 10:33

Page 7

fig Hu-Unc53/1 seq (1 > 6013) Site at sequence

TCATGAGCGTGACCGCGGGCGGGTSSACCTCGCCCTGTGAGCAGCAGCATTGACCCCAAGTCCTCAACCAAGCAGGGAGGCTTACGCCTTCCA
AGATACTCGCAC TGGCCGCCCGCCCCACCTGGAGCGGACAC TCGTCGTCGTAAC TGGGGTCAGAGGAGTCGTGGTTCTGTCCCTCCGGAATGCGGAAGG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53rv4

S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GAC TGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCGTGCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC
CTGACTTCCTCGGATGGTTCCATCGGTACCCGCCCTGGTGAGGTGGGGACAGTTAGTCTGTCTAGCCCCTTTCTCTTCCGGTTTCGGTTCCGTCACCG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A V A

CTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTG
GAACCTGAGTCTGTTGTAGAGGAACCTTCATAACCGAGGGGCTTTTCATGAGGGTTCTTGGTTTCGTTTCGGTGGGGTGTCTGGTGGTTTCGACCGTCTCGAC

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S : G S P E S T P K N Q A S H P T A T L A E L

CCAGCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTS TCAAAACACCC TCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCAT
GGTGGTTGGGGAGAGTCCCGGTGTGCTTCTCGAAACAGTTTGGTGGGAGTGCATGGTTAGAACGTTTCCAGTTGAGGTGTGTCAGACCTAGATGGTAGTA

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L P S

Fig 9

Tuesday, 18 November 1997 10:33

fig HU-Unc53/1 seq (1 > 6013) Site and sequence

Page 4

GGTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTTCGAAGTCAGGGTCCCTCTCTGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGG 250

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L F

TCTCCCCCTGCACCTTCCCATGTCTCTGAGTGCAGAGGGCAACTTACCAACATAGTGAGTCCCACCTGCGGCCACCAAGCAAGCAATCACCCGCTCCAACA
AGAGGGGGACGTGAAGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGCGGCAGGTTGT 260

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGCGGCCCTCGAGCTGACAGCGGCTCCCAATGGGGAGCACCTGTCCCTGGCCGAGAGACCAAGGGCAATGATTGGGTCAAG
CGTAGGGGTGGTGCTCCGCCGAAGCTCGACAAGTCCCGGAGGGTTTACCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC 270

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L S G S O M G S T L S L A E R P K G M I R S G

ATCCTTCGAGAGCCCCACGGACGATGTTACGCGCTCACTGCTGCTCCCTGGCCTCCAGTGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATGT
TAGGAAGGCTCTGGGGTGCTGCTACAAAGTGCAGTCAAGCAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCTTACGTTAGA 280

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T G D V H S S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAAAATCCGGAAGCTTCGTAGGGAAGTGGAAATCATCCAGGAAAAAGTGGCCACCTTGACGCTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCT
CTCGTTTAGGCTTCGAAGCATCCCTTGACCTTATAGGCTCTCTTACCCGGTGGAACTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA 290

U2 ORF = pCB351 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P K L F P E L E S V E K V A T L T S O L S A N A N L V A A

Tuesday, 18 November 1997 10:33
fig HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page

TTGAGCAGAGCC TGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGC TGGATTTCGAGAAACCATAGA
AACTCGTCTCGGACCACCTATAC TGTAGGGCGGACGC TG TGGACCGTC TC TCCCGGC TCCTC TTCC TGTGACTCGACGACCTAAACGCTCTT TGGTATCT

300

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N N T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCGAGAGAAAAAGAACTCTGAGGCCAGGCACTCATTCAGGGAGCCCTTAATGCC TCAGAAACACACCCAAAGAACTTCGGATCAAGAGACAAAG
GAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCC TAGTCTCTGTTTTG

310

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q H

TCC TCAGATAGCATCTCAAGCCTCAACAGCATCAC TAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCTTCTACGACTACGCTTTTCTCTTTTTTCTCAACCC

320

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S ! T S H S S I G S S K D A D A K K K K K S V

TCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGTTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATGTGGGT

330

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S O I E E I A T P D

CTCTTCAGCGCCCTCATCCCGCAAACTACAGCATGCTTACAGAGACTGCTTACCGTCCATCAAGTCTCCACCTTGTCTCCGTGGGCACTGATGT
GAGAAGTCCGGGGAGTAGGGGGTTTGTATGTCTACCAAGATGCTCTGACGAGGTGGGAGGTAGTTTCAGGAGGTGGAAACAGGAGGCAACCGTGACTACAG

340

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

I S A P S S P K L Q - S S T E T A S P S I K S S T L S S V S T D V

Tuesday, 18 November 1997 10:33
fig. Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 8

ACCG .GCCCTGCTCAGCCAGCCCCCACATAGGCTGTTCATGCAATGAGGAGGAGGAGCCAGAGAGAAGAGGAGGTATCGGAGCTGGCTCTGAGG
TGGCTCCCGGACGAGTGGGTGGGGGGGTGATCCGACAAGGTACGTTACTCTCTCTCTCGGTCCTTCTTCTTCCATAGCCTCGACGCGAGACTCG 3600

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAAATGCAGT
ATACCTCTTCTTCTTACTTCGAATGTCTGTAGGCGAACCCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAA 3600

U2 ORF = pCB251 ORF

PHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATT
CCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAA 3700

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCCACGCCGCTCCCTAGGCCTGGCACTGACCCATTCCTTCGGCCCCAGTCTTGACAGACAGACCTGTCAACCATGAGTGGCACTGACTTGTG
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGATGGGTAAGGAAGCCGGGTCAGAACGTCTGTGTC TGGACAGTGGGTACCTACCGTAGTCATGAACAC 3800

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L - S F G P S L A D T C L S P H D G : S T C

Tuesday, 18 November 1997 10:34
fig. 'u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 9

GTCCAAAGGAGGAAGTGACCCGCCGGGTGGTGGTGAAGATGCCCGCCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTTCTTCTGGGCTGTAG
CAGGTTTCTCTCTTCACTGGGAGGCCACCACCACCTCTACGGGGCGTCGTGTAGTAGTTTCCCTGAACTTCGTCTCTTAAAGAGGACCCGACATC

391

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAATGGACCCAGCCTCTACCTGGGA
GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCCTTGATATAAAGATTTTACC TGGGTCGGAGATGGGACCT

400

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

GTAAGCACTGAGTCCATCCATGGCTACAGCAACAGCCACGTGAACGAGTGTTGGATGCGAGAGCCCCGAGATGCCTCCTTGCCGTGAGGTGTCATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTGGTGCACCTTTGCTCACAACCTACGTCTCGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT

410

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R S G V N

AGATATCAGTCTCCCTCAAAGGCTGAAAGGAGAAATGGTGGACAGCCTGGTGTTCGAGAGCGCTGATCCCAAGCCGATGATGCAGCACTACATAAGGCT
TGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTACGACGCTGTCGACCAACAAGCTCTCGGACTAGGGTTCGGCTACTAGTCTCTGATGATTTGGGA

420

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

H I V S L G L E C D S L V F E T L P P P M V Q - I D L

Fig. 9

Tuesday, 18 November 1997 10:34

Page 10

lig lu-Unc53/1 seq (1 > 6013) Site and Sequence

CCCTCTGAAGCACCGGCGCCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAA*CGCTTGGCCGAGTACC TGSTGGAGCGCTCTGGG
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTCGCCGTGCCCGTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCCAGATAGACC
GCACCTCCAGTGCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E V T E S I V S T F N M H Q Q S C K D L Q L Y L S N L A H Q I D

GGGAAACAGGAATTGGGATGTGCCCTCGTGATTCTATTGGATGACCTGAGTGAAGCAGGC TCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA
GGCTTTGTGCTTTAACCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCGAGGAGTCACTCAACCACTTACCCGGGAGTGGAGCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E T S I S D V P L L L C D L S E A G S I S E L V V G A L T C I

Tuesday, 18 November 1997 10:34

Fig 9

Page 11

fig 4u-Unc53/1 seq (1>6013) Site Sequence

GTATCAAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCAACCATGGCTTGCACCTGAGCTTCAGGATGTTGACCTTCTCC
CATAGTATTTACAGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAAGCTGAACTCGAAGTCTTCAACTGGAAGAGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72626H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCCTGGTTCTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC
TTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGCTGTGCTGTAGTTACGGTTGTTCCTTCTCGACGAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGTGTCTGAGCTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCCTTGAAGACACAGCACCTCAGACTTCTCATCGGCCCTTGCTTCTTCTGTG
CCCACGAGCTGACCCATGGTTTGACACCATAGTAGAGGTGTGGAAGGAACCTTCTGTCTGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L S A V P K L A Y L H T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34
fig. 1J-Unc53/1 seq (1 > 6013) Site an. sequence

Fig. 9

age 12

GTSTCCCATTTGGCATTGAGGACTTCGGGACCTGGTTCATTGACCTGTGGAAACAACCTATCATTCCTCTATCTACAGGAAGGAGCCAAAGGATGGGATAAAG
CAGAGGGTAACCGTAACCTCTGAAGGCCCTGGACCAAGTAACCTGGACACCTTGTGAGATAGTAAGGGATAGATGTCTCTCTCGTTCTCTACCTATCTCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L O E G A K D G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCCTGCECCATCAGCCCAACAAGACCAATCAAAGCTGTACCAAC
CAGGTACCTGTCTTTCGACGAACCTCTCTGGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTGTCTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V - S C K A A V E D P V E V V R D T L P W P S A Q Q D Q S K L Y H

TCCGCCCAACCCACCGTGGGCCCTCAGAGCATGGCTCACCTCCCGAGGATAGGACAGTCAAGSACAGCACCCCAAGTTCTCTGGACTCAGATCTCTGA
TCCGGGGTGGGTGGCACCCTGGGAGTGTCTGAACGGAGTGGAGGGCTCCTATCTCTGTCTGTCTGGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L P P P T V G P - S I A S P P E D R T V K D S T P S S L D S D - L I

Tuesday, 18 November 1997 10:34

fig	tu-Unc53/1 seq (1 > 6013)	Site a	Sequence
-----	---------------------------	--------	----------

Fig. 9

Page 13

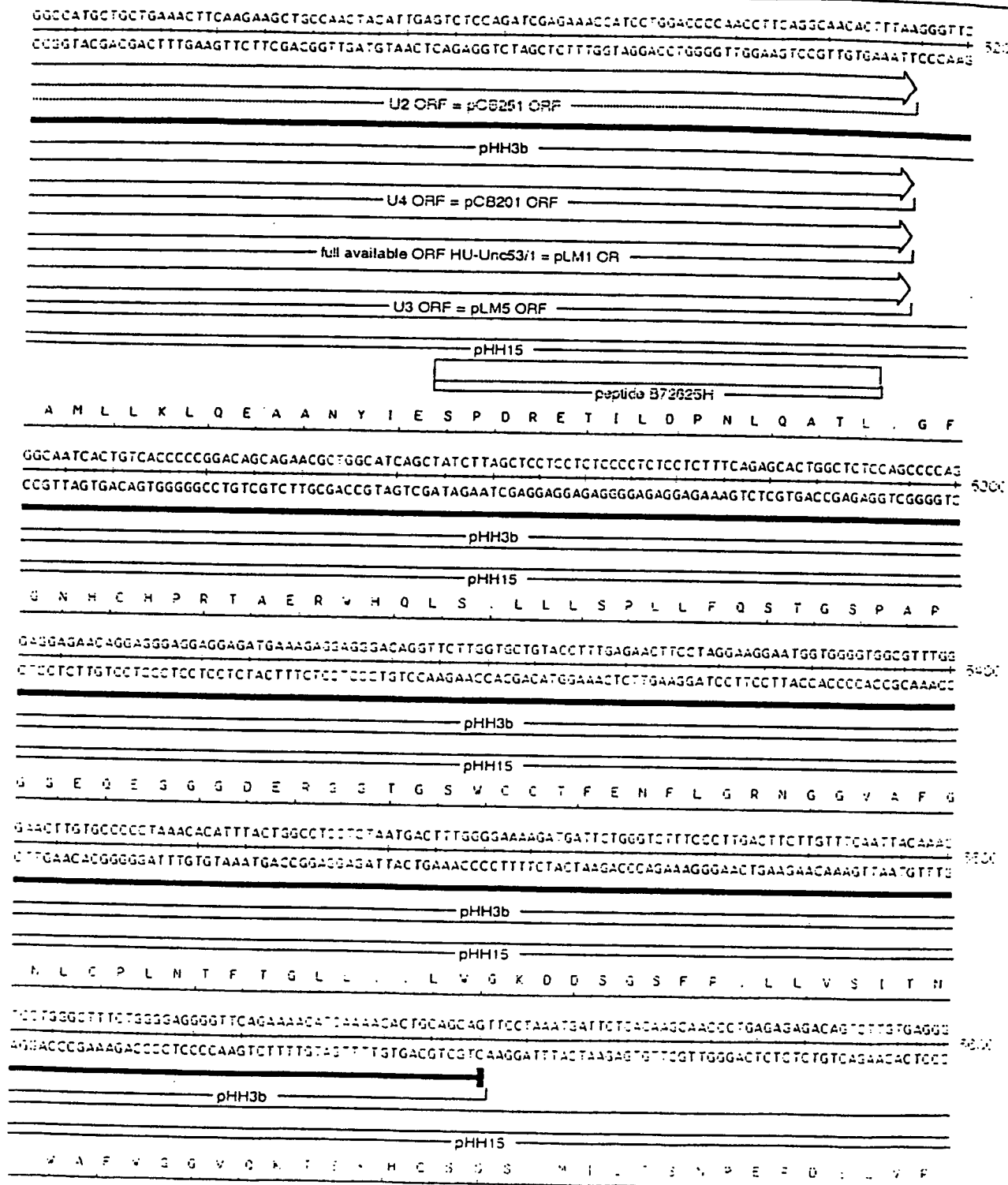


Fig 9.

Tuesday, 18 November 1997 10:34

fig 1Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 17

AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCACAGACCCTTCCCAATTCATCACCCTGCCAACAACCTCTCCCCAGAGATCTGGCTGAGGC
TCTAGACCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGCTGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGCTCTAGACCGACCTCTG 570

————— pHH15 —————>

E I W G R Q E A P Q I F S Q T L P N S I T T A N N S S P R D L A G A

CCAGAAAAAGAAGCATGTGGTTTAAAAATGTTAAATCAATCTGTAAAAGGTAAAAATGAAAAACAAAAACAAGCAACAACAAAAACAATGGAAA 580

GGTCTTTTCTTCGTACACCAAAATTTTACAAATTTAGTTAGACATTTTCCATTTTACTTTTTGTTTTGTTCTGTTTGTGTTTTGTTTACCTTT 590

Q K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q V I

AGATGAAGCTGGAGAGAGAGGAACCAAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAAC 590

TCTACTTCGACCTCTCTCTCTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCTGTAGTTGTTCTGCCGACGGTTGG 600

R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A N

TGAGAAGTCACCAAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAG 600

ACTCTTCAGTGGTTTGGTGTTTTATTGGAATGTCGGAAGTCCCTTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC

linker? —

L R S H Q T T K I T L Q P S G K D Y Q L C L S T L . F N N A P E F S

CTTGGACTTAACC 6013

GAACCTGAATTGG

————— linker? —————>

L D L T

[illegible][illegible]

AGTC TCCCTC AAGGTC GAGGAG AATGCG TCGACG CCGTGG TGTTCGAG CCGTGA TCCCAAG CCGATGATG CAGCCTACATAG CCGTCC TCGTG 1900
V S L K G L K C K C V O S L V F E T L I P K P A A O H V I S L L L
homology block E - pred. nucleotide 80

AAGCAC CGGCGCC TCGTCTCTC GGC CCGAGGGCAGGGCAGACCTACCTGACCAATCGCTTGGCCGAGTACC TGGTGGAGCGCTCTGGCCGCTGAGG 2000
K H R R L V L S S P S G T G K T V L T H A L A E Y L V E R S G R E
homology block E - pred. nucleotide 80

TCACAGAGGGCATCGTACGACCTTCAACATGCGCCAGCAGTCTTGCAGGATCTGCAATCTGATCTTTCACACCTAGCCACAGATAGACCGGGAAAC 2100
V T E G I V S T F N A H O D S C K O L O L V L S N L A N Q I O R E T
homology block E - pred. nucleotide 80

AGGAATGCGGATGTGCCCC TCGTAT TCTAT TGGATGACC TGAGTGAGCAGGCTCCATCAGTGAG TTGGTCAATGGGCCCC TCACC TGCAAGTATCAT 2200
G I G O V P L V I L L O D L S E A G S I S E L V N G A L T C K Y H
homology block E - pred. nucleotide 80

AAATGTCCTTATTTATAGG TACCAACATCGCCTGTAAATATGACCCACCACTGGCTTGCACCTTGAGC TTCAAGATGTTGACCTTC TCACACACG 2300
K C P V I I G T T N O P V K A T P N H G L H L S F R A L T F S N H
homology block E - pred. nucleotide 80

TGGAGC CAGCCATGGCTTCC TGGTTCGTTACCTGAGGAGGAGCTGG TAGAG TCGACAGCGACATCAATGCCAACAGGAGAGCTGCTTCGGGTGCT 2400
V E P A N G F L V R Y L A R K L V E S O S O I N A N K E E L L A V L
homology block E - pred. nucleotide 80

CGAC TGGGTACCCAGGCTGTGGTATCATCTCCACACCTTCC TTGAGAGCAGCAGCACC TCAGACTTCC TCATCGGCCCTTGCTTCTTTC TGTGCTGCTCC 2500
O V V P K L V V H L H T F L E K H S T S O F L I G P C F F L S C P
homology block E - pred. nucleotide 80

ATTGGCATGAGGACTTCCGACCTGG TCTATGACCTG TGGACACAC TCTATCATTCCTTATCTACAGGAGGAGCCAGGATGGGATAAAGGTCCATG 2600
I G I E O F R T V F I O L V N N S I P V L O E G A K O G I K V H
homology block E - pred. nucleotide 80

GACGAAAGCTTCTTGGAGGACCGAG TGGATGGGTCCGGACACAC TCCCTGGCCATCAGCCACACAGACCAATCAAGCTGTACCACTTCCCCC 2700
G O K A A V E O P V E V R O T L P V P S A O O O S K L Y H L P P
homology block E - pred. nucleotide 80

ACCCACCGTGGGCCCTCAGCAGCTTCCCTACCTCCCGAGGATAGGACAGTCAAGACAGCAGCCCAAGTTC TCGGACTCAGATCTCTGTGGCCATG 2800
P T V G P H S I A S P P E O R T V K O S T P S S L O S O P L N A H
homology block E - pred. nucleotide 80

CTGC TGAACCTTCAAGAGC TCCACAC TCAATGAGTCTCCAGATCGAGAACCATCC TGGACCCCAACCTTCAGGCACACTTAAAGGGTTCGGCACTC 2900
L L K L O E R A N Y I E S P O R E T I L O P M L Q A T L
homology block E - pred. nucleotide 80

AC TG TACCCCECCAGCAGCAGACGCTGGCATCAGATATCTTAGCTCC TCC TCCCTC TCTCAGAGCAC TGGCTCTCCAGCCCEAGGAGGAGA 3000
3' untranslated trailer

ACAGGAGGGAGGAGGAGATGAAATAGGAGGAGGATTC TTGGTGC TGTACCTTTGAGACATTC TAGGAGGAGATGGTGGGG TGGCGTTTGGGAC TTG 3100
3' untranslated trailer

TGCCCCCTAAACACATTTAC TGGCTCCCTTATGAC TT TGGGAAAGATGATTC TGGCTTTCCTT TGACTTC TTGTTTCAAT TACAACTCTCTGGG 3200
3' untranslated trailer

CTTCTTGGGAGGGGTTCAGAAACATCAAAACATGCGCAGTTCCTAATGAT TCTCAGACACACCTGAGAGAGCAGCTCTTGTGAGGAGATCTG 3300
3' untranslated trailer

GGGAGGAGGAGGAGCTCTCAGATTTCTCAGACACCTTCCATTCCTCAGCAGTGCACACAC TCC TCCCCAGAGATCTGGCTGGAGCCAGAA 3400
3' untranslated trailer

AAAGAGCATGTGGTTTAAAAATGTTTAAATTAATCTGTAAAGGTAAATTAAGAAACAAACCAAGCAACAAACAAACCAATGGAAAGATGA 3500
3' untranslated trailer

GC TGGAGAGAGTGTACCTGTTGGTAAATAGAGC TGGCCCTCTGCTGCTTGGATGATAGGAGACATCAACAAACAGGCTGCCACCTGAGAGG 3600
3' untranslated trailer

TCACCAACACCAAAATATCTTACCTCTTGGGAAAGCTACAGCTCTGCTTCTACCTCTTATTTACCAATGACCGGAAATTCAGCTTGGAC 3700
3' untranslated trailer

TTAACC 3706



Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 2

GCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGTGAACGGGGGTTCTGTCACACAGC 1500
CGGTACCCGCTATTTCAGCACAGAATGGCCCAACCTGAGTCTGCTATCAATGGCCATTTCGCGTCCGACGCCGACTTCCCCCAAGCACGTGTGTCG
A S G D K S C L T G L D S R R . L P D K A Q R S G . T G G S C T Q

CCAGCTTGGAGCGAACGACCTACACCGAATGAGATACCTACAGCTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGCGGACAGGTATCC 1800
GCTCGAACCTCGCTTGTGATGTGGCTTGAATGTGCGACTCGATACCTTTTCGCGGTGCGAAGGGCTTCCCTCTTTCGCGCTGTCATAGG
P S L E R T T Y T E L R Y L O R E L . E S A T L P E G R K A D R Y P

GGTAAGCGGCAGGGTCGGAACAGGAGCGCAGGAGGAGCTTCCAGGGGAAACGCCCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTCTGACTT 1700
CCATTTCGCGTCCCGAGCTTGTCTCTCGCGTGTCTCCGAGGTCGCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAGCGGTGGAGACTGAA
V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L . L

GAGCTGCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAACACCGCAGCAACGCGGCTTTTACGGTTCCTGGCTTTTGTGCGCTTTTG 1800
CTCGCAGCTAAAAACACTACGAGCAGTCCCGCGCTCGGATACCTTTTTCGCGTGTGCGCGGAAATGCCAAGGACCGGAAACGACCGGAAAC
E R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F

CTCACATGTTCTTCTCGCTTATCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGATACCGCTCGCGCAGCCGAACGACCGAGCG 1900
GAGTGATACAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTGACTATGGCGAGCGGCTCGGCTTGTGCTGCG
A H M F F P A L S P D S V D N R I T A F E . A D T A R R S R T T E R

CAGCGAGTCAGTGAGCGAGGAAGCGGAGCGCCCAATACGCAACCGCTCTCCCGCGCTTGGCGGATTCATTAATGACGTGGCAGCAGAGGTTT 2000
GTGCTCAGTCACTCGCTCTTCCGCTTCTCGCGGTATGCGTTTGGCGGAGAGGGGCGCAACCGGCTAAGTAATACGTCGACCGTGTGTCGAA
S E S V S E E A E E R P I R K P P L P A R V P I H . C S V H D R F

CCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTAGTACCTCACTATTAGGCAACCGGCTTTACACTTTATGCTTCCGGCTCGTATGT 2100
GGCTGACCTTTTCGCGCTCACTCGGTTGGCTTAATACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCGGAGCATACA
P D V K A G S E R N A I N V S . L T H . A P Q A L H F H L P A R M

TGTGTGGAATGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATTACGCCAAGCGGCAATTAACCTCACTAAAGGAAACAAAGCT 2200
ACACACCTTAACACTCGCTATTGTTAAAGTGTGCTTTTGTGSA"ACTGGTACTAATGCGGTTCGCGGTTAATGGGAGTGATTCCCTTTTTCGA
L C G I V S G . O F H T G N S Y D H D Y A K R A I N P H . R E O K L

GGGTACCGGCGCCCTCGAGGTGAGCTATCGATAACTTGAATTCGCAATTCCTGACGCCCTGCTTTCAGCCAGATGCTGGACCCAGAGTCCCGAG 2300
CCCATGCGCGCGGGGAGCTCCAGCTGCCATAGCTATTCTCACTA"AGCTTAAGGACGTCGGGGACGAGAAGTCGCTACGACCTGGGTCTCAGGGTCT

insert pLM1

ORF pLM1

G T G P P L E V D G I D < L D ! E F L Q P L L F S Q M L D P E S Q

AGAAAGAGGACAGTSCAGAATGCTCTGGATCTCGGCGAGAACTTGGAAGAGACCATGTCCAGCTGCGAGGGTCCAGGTGACTCAGCTCCCTGGAGA 2400
TCCTTCTCTGTCAGCTCTTACAGGACCTAGAGGCGCTTGGACCTCTCTGGTACAGGTGGGACGCTCCAGGGTCCACTGAGTGTCGAGGGACCTCT

insert pLM1

ORF pLM1

R < R T V Q N V L D L R Q N L E E T M S S L R G S Q V T H S S L E

TAACCTGCTACGACAGCGATGATGCCAACCCAGCAGCTST"CACTCTTCCAACCGCTGCTCCCTCTGTCATGGCGCTATGGCAGTCCAGTCCGCG 2500
ATTGGACGATGCTGTGCTTACTACGGTTGGTSCGTGCAAGGCTTGGAGAGTGGGAGCAGGGAGACAGTACCGGATACCGGTCAGGTCAGGCGCT

insert pLM1

ORF pLM1

M T C Y D S D A N P R S V S S L S N R S S P L S V R Y G S S P R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 3

GC TGCAGGCTGGTGACGCCCTCTGTGGGTGGGAGCTGCTCTGAGGGGACGCCCGCTTGGTACATGCACGGGGAACGGGCCAC TACTCCACACC 2800
CSACGTCCGACCAC TGGCGGGAGACACCCACCTCGACGGGAGGCTCCCTGCGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTG TGG
-----insert pLM1-----
-----ORF pLM1-----
L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H T
ATGCCCATGCGCAGCCCCAGCAAGCTCAGGCATATCTCCCTGCTGGAGCTGGTGCAGATCCCTGGACTCGGATGAGGTGGACC TCAGTCCGGCTACATGA 2700
TACGGGTACGGTCGGGGTCTGTTCGAGTCGGTATAGAGGCGGACCTCGACCAGCTTAGGGACCTGAGCCTACTCCACCTGGAGTTCAGGCGGATGTACT
-----insert pLM1-----
-----ORF pLM1-----
M P H R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M
GCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACCCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCCTC 2800
CGCTGTCACTGGAGTACCGCTTCTGGTACTGCTCTACTACTGTAGTATGGGCGGACCTACTTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAG
-----insert pLM1-----
-----ORF pLM1-----
S D S D L M G K T M T E D C D I T T G V D E S S S I S S G L S D A S
AGACAATCTCAGTTCAGAAGAATCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCATGCTTCTCGCAGGAACCAACAATAGTGTACGC 2900
TCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAATGAGGGAGGGTTCATGAGGGTGACGAAGAGCGTCTTGAGTTGTTATCACGATCGG
-----insert pLM1-----
-----ORF pLM1-----
D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L R
ACAGACTCAGAGAAGCGCTCAGTGGCAGAAAGTGGGCTGAGCTGGTTAGTGAATCAGAGGAGAAAGCCCTAAAAAATGGAGTACGACAGTGGTAGCC 3000
TGCTTGAGTCTCTTCGCGAGTGACCGCTTTTACCCGAGCTGAGCAAAATCACTTAGTCTCTCTTTTCGGGGATTTTTGACCTCATGCTGTCACCATCGG
-----insert pLM1-----
-----ORF pLM1-----
T D S E K R S L A E S G L E V F S E S E E K A P K K L E Y D S G S
TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAAGCTGAGCTGAGTGTATGATTCATCCAAGGGTGGAGAATGAAAAAGCCCATCAGCCTGGG 3100
ACTTCTACCTTGGACCCCTGAAGATTACCGGCTCCCTGAGCTGAGACACTAAGTAGGTTCACCTCTTGACTTTTTCGGGTAGTCGGACCC
-----insert pLM1-----
-----ORF pLM1-----
L K M E P G T S K V R R E P E S C D S S K G G E L K K P I S L G
GCAAGCTGGTTCCTGAAGAAGGGCAAGACCCACCTGCTGCTGAGCTTCCCATCACTCAGACAGCCAGAGTGCCTCAAGTTCGACGGCAACCT 3200
GGTGGGACCAAGGAGCTTCTTCCGCTCTGAGGTGAGTACCTGAGTGAAGGGGTAGTGAGTGTGCGGGTCTACGGGAGTTTCAGCGTCCGTTTGA
-----insert pLM1-----
-----ORF pLM1-----
H P G S L K K G K T P P V L V T S P I T H T A O S A L K V A G K P
GAGGGCAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGTATCTGCTGCAACGCTCCTCTCTGATGCTGGTGGGACCGCTGAGTGTGCTAAGA 3300
CTGCGCTTTCGATGCTGTGTCCTCATTCGACCTCACTTCTTCTGAGTGGTGGGAGGAGACTACGACCAAGCCCTGGCGGACTCACTACGATCTCT
-----insert pLM1-----
-----ORF pLM1-----
E I K A T D K G K L A V K L L D R S S S D A G R C P L S D A K

Page 4

BNSDOCID: <WO 9824810A2 | >

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 5

CAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAGAGACCCGATGTACCCCAAACCTCAGGCC TGCACAGGAGCATGGAGTCCCTCCAGATGC 4200
GTCCCGGACCTCGATTACACCAAAGTCACACGGTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTGCTCGTACCTCAGGGAGGTCTACG
-----insert pLM1-----
-----ORF pLM1-----
Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L O M
CAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCCACTGCTCCCCCTGCTGCTCCACAGAAGAAGAGACGGAAGAGCTGACTTG 4300
GTTACTCGGAGGGGTCACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGTGCTCTTCTCTGCCTTCTCGACTGAAC
-----insert pLM1-----
-----ORF pLM1-----
P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T V
GAGTGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGA TCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAG 4400
CTCACCTTGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTCTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCCTCCCT
-----insert pLM1-----
-----ORF pLM1-----
S G S P R A G Q L D S N Q R D R N T L P K K G L R Y Q L Q S Q E E
ACCAAGGAGAGGGCAGATTCACATACCATTTGGTGGGCTGCCGAA TCCGATGACCACTCAGAGCTGCCCTTCTCCCCCTGCCTTCCCATGCTCTGAGTG 4500
TGGTTCCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTT AGGC TACTGGTCACTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCAC
-----insert pLM1-----
-----ORF pLM1-----
T K E R R H S H T I G G L P E S D D Q S E L P S P P A L P M S L S
CAAAGGGCAACTTACCAACATAGTGAAGTCCCACTGCGGCCAGCA TCCCAAGAATCACCCGCTCCCAACAGCATCCCCACCCAGGCGGGCTTCGAGCT 4600
GTTTCCCGGTTGAATGGTGTATCACTCAGGGTGAAGCGCGTGGT TCGGTTCTTAGTGGCGAGGTGTCTGAGGGGTGGGTGCTCCSCCGGAAGCTCGA
-----insert pLM1-----
-----ORF pLM1-----
A < G O L T N I V S P T A A T T P R I T R S N S I P T H E A A F E L
GTACAGCGGC TCCCAATGGGAGCACCTGTCCCTGCGCGAGAGAC CCAAGGGAATGATTCGGTCAAGATCCTTCCGAGACCCACGAGCATGTTTCAC 4700
CATGTGCGCGAGGGTTTACCCCTCGTGGGACAGGGACCTGGTCT TCGGTTCCCTTACTAAGCCAGTCTTAGGAAGGCTCTGGGGTGCCTGTACAAAGTG
-----insert pLM1-----
-----ORF pLM1-----
* S G S Q M G S T L S L A E P P K G M I R S G S F R D P T D D V H
GAGTCAGTGC TGCCTGGGCTCCAGTGCCTGCTCCAGTACTGCT CAGCTGAGGAGAGGATGCAATCTGAGCAAA TCCGGAAGCTTCGTAGGGAAGCTGG 4800
CTGAGTCAAGACAGGGAGCGGAGGTCACGGAGGAGGTTGATGAG GATGAGGATGACTCTCTCTACGTTAGACTCGT TTAGGCTTCGAAGCATCCTTTGACC
-----insert pLM1-----
-----ORF pLM1-----
G G V L S L A S S A S S T S S A E E R M O S E Q I R K L R P E L
AATCATCCCAAGGAAAAAGTGGCACCTTGACGTCTCAAGTTCTG TCTTAATGCTAATCTGGTGGCTGCTTTTGGACAGAGCTTGTGAATATTACATCCCG 4900
TTAGTAGGGTCTTTTTCACCGGTGGAACTGCAGAGTTGAAGGAT TACGATTAGACCAACGACGAAAACTGCTCTGAGCACCTTATATGTAGGGC
-----insert pLM1-----
-----ORF pLM1-----
S G S E A V A T L T S S L S A N L V A A F E Q C L V Y M T S R

Page 6

BNSDOCID: <WO 9824810A2 | >

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

AAGGACAGGTTCTTGGTGTGACCTTTGAGAACTTCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTGTGCCCTAAACACATTTACTGGC
TCCTCCCTGTCCAAGAACACGACATGGAACTCTTGAAGGATCCTTCTTACCACCCACCCGAAACCTTGAACACGGGGATTTGTGTAATGACCG 7400

Insert pLM1
G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G

CTCTCTAATGACTTTGGGAAAAGATGATTCTGGGTCTTCCCTTGACTTCTTGTTCATTACAACTCTTGGGCTTTCTGGGGAGGGGTTGAGAAA
GAGGAGATTAAGAACCCCTTTTCTACTAAGACCCAAGAGGGAACGAAGAACAAGTTAATGTTGAGGACCCGAAAGACCCCTCCCAAGTCTTTT 7500

Insert pLM1
L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K

CATCAAAACACTGACGAGTTCCCGGAATTCAGCTTGGCTTAACGAGGCTGAACCTTGCTCAAAAGAGCGAATTCAGCACACTGGCGGCCGTACT
GTAGTTTGTGACGTCGTCAAGGGGCCCTAAGTCGAACCTGAATTTGGTCCGACTTGAACGAGTTTCTTGGCTTAAGTCGTGTGACCGCCGGCAATGA 7600

Insert pLM1
T S K H C S S S P E F S L D L T R L N L L K R S R I P A H V R P L L

AGTTCTAGAGCGGCCGCCACCGCGGTGGAGCTCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCGTCTTTTACAACGTCGTGACTGG
TCAAGATCTCGCGCGCGGTGGCGCCACCTCGAGGTAAAGCGGATATCACTCAGCATAATGCGCGCGAGTGACCGGCAGCAAAATGTTGACGACTGACC 7700

→
V L E R P P P R V S S N S P Y S E S Y A R S L A V V L Q R R D V

GAAAAACCTTGGCGTTACCCAACCTTAATCGCCTTGACGACATCCCCCTTTCCGCGAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAAC
CTTTGGGACCGCAATGGGTGAATTAGCGGAACGCTGTAGGGGAAAGCGGTGACGCGCATTATCGCTTCTCGGGCGTGGCTAGCGGGAAGGGTTG 7800

E N P G V T Q L N R L A A H P P F A S V R N S E E A R T O R P S Q

AGTTGCGCAGCTGAATGGCGAATGGGACGCGCCCTGTAGCGCGCATTAAAGCGCGCGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
TCAACGCGTGGGACTTACCGCTTACCTGCGCGGACATCGCGCGTAAATCGCGCGCCACACCAATGCGCGTGGCGATGGAACGGTC 7900

Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S

CGCCCTAGCGCGCGCTCTTTGGCTTTCTTCCCTTCTTCTGCGCACGTTGCGCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTT
GCGGATCGCGCGCGGAGGAAAGCGAAAGAGGGAAGGAAAGAGCGGTGCAAGCGCGCGGAGGAGGAGGAGTTGAGATTAGCCCCGAGGGAATCCCAAG 8000

A L A P A P F A F F P S F L A T F A G F P R Q A L N R G L P L G F

CGATTTAGTGCTTTACGGCAGCTCGACCCCAAAAACCTTATAGGGTGATGGTTACGTTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCCCTTTGA
GTAAATCACGAAATGCGGTGGAGCTGGGGTTTTTGAACCTAACTCCACTACCAAGTGATCACCGGTAGCGGGACTATCTGCCAAAAGCGGGAACCT 8100

R F S A L R H L D P K K L D . G D G S R S G P S P . . T V F R P L

CGTTGGAGTCCAGCTTCTTTAATAGTGGACCTTGTGTTTGAACCTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGC
GCAACCTCAGGTGCAAGAAATATCACTGAGAACAAGTTTGAACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAACCTAAATATTCCTTAAACGG 8200

T L E S T F F N S G L L F Q T G T T L N P I S V Y S F O L . G I L P

327TTCGGCTATTGGTTAAAAATGAGCTGATTTAAACAAAAATTAACGGAATTTAAACAAATATTAACGCTTACAATTTAG
TAAAGCCGATACCAATTTTCTACTGACTAAATTTTAAATTCGCTTAAATTTGTTTATAATTCGCAATGTAAATC 3285

S A Y V L K N E L I . S K F N A N F N K I L T L T I .

SEQUENCE IDPCT

Tuesday, 18 November 1997 11:48

fig 34 pLM4 (1 > 10070) Site and Sequence

Enzymes: 100 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

16p

100
TAGTTATTAATAGTAATCAATACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCC
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGCGGACCGACTGGC

200
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTACGGT
GGGTGGCTGGGGGCGGGTAAGTGCAGTTTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTCATAAATGCCA

300
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAT

400
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCAC TACGCCAAAACCGTCATGTAGTTACCCGACCT

500
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V
TAGCGGTTTGACTCACGGGGATTCCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGGTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCTAAAGGTTGAGAGTGGGGTAAGTGCAGTTACCTCAAACAAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT

600
I A V . L T G I S K S P P H . R Q W E F V L A P K S T G L S K M S
ACAAC TCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTTGTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAAGTGGCTTACCCGCCATCCGCACATGCCACCTCCAGATATATTCGTCTCGACCAATCACTTGGCAGTCTAGCGCATCGCGAT

700
D L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCSAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTGGCGGTGTCAAGTGGC

800
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCGCGGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGAC
ACAGGCGCGCTCCCGCTCCCGCTACGGTGGATGGCGTGGTACTGGGACTTCAAGTAGACGTGGTGGCGGTTCGACGGGCACGGGACCGGTGGGAGCACTG

900
V S G E G E G D A T Y S K L T L K F I C T T G K L P V P W P T L V T
CAGCGTGACCTACGGCGTGAGTGCTTCAGCGCTACCGGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTGGCGCATGGCGCTGGTGTACTTCGTGCTGCTGAAGAAGTTCAGGCGGTACGGGCTTCGGATGCCAGTCCCTC

L T Y G V G C F S R Y P D H N K Q H D F F K S A M P E G Y V D E

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 2

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCC TGCTGCCGTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCAC TTGGCGTAGCTCGACTTCCCGTAGC 100

EGFP

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCCTGCCGTGTAGGACCCGTTGACCTCATGTTGATGTTGTCGGTGTGTCAGATATAGTACCGGC GTTTCGTCTTCTGCCGTAGTT 110

EGFP

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I A

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCCTGCCGTGCACGCTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 120

EGFP

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGTGAGGTTCTGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTGAGGCGGGACTCGTTTCTGGGGTGTCTCTCGCGCTAGTGATACCAGGACGACCTCAAGCACTGGCGGGCGGCCCT 130

EGFP

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGACTCAGATCTCGAGCTCAAGCTTCGAATTCGAGTTCGATAAGCTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGCTTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTGAACTATAGCTTAAGGACGTCGG 140

EGFP

I T L G M D E L Y K S S L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAAGCAGATGCTGGACCCAGAGTCCAGAGAGAGGACAGTGCAGAATGTCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGCTCTTTCTCTGTCACGCTTACAGGACCTAGAGGCCGCTCTGGACCTTCTCTGGTACAGGTCTG 150

Insert pLM1

ORF pLM1

L L F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S

CTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCGAGTGTCCAGCCTCTCCAACCGCTCG
GACGCTCCAGGGTCCACTGAGTGTGAGGGACCTCTACTGGACGATGCTGCTGCTACTACGGTTGGGTGCGTCCGACAGGTCGGAGAGGTGGCGAGGA 160

Insert pLM1

ORF pLM1

L R G S Q V T H S S L E M T C Y D S D A N P R S V S S L S N R S

CCCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGACAGGTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCTG
GGGAGACAGTACCGCGATACCGGTTCAGGTGAGGCGGCTGACGTCGACCACTGCGCGGGAGACACCCACCTCGACGGGAGGCTCCCCGCGGGGAGAC 170

Insert pLM1

ORF pLM1

L P L S R / G S S S P R L G A G D A P S V G G S C R S E G T P A W

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 3

GTACATGCACGGCGAACGGGCCCCACTACTCCACACCATGCCCATGCGCAGCCCCASCAAGC TCAGCCATATC TCCCGCC TGGAGCTGGTGGAAATCCCTG
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCTGTCGAGTCGGTATAGAGGGCGGACC TCGACCAGCTTAGGGAC

insert pLM1

ORF pLM1

Y M H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L

GACTCGGATGAGGTGGACC TCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG
CTGAGCC TACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGCGCGACCTAC

insert pLM1

ORF pLM1

C S D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D

AAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCC TCAGACAATCTCAGTTTCAAGAATTCATGCCAGCTCCTCACTCAAC TCCCTCCCAAGTACTCC
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGG

insert pLM1

ORF pLM1

E S S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P

CAGTGCTTCTCGCAGGAAC TCAACAATAGTGCTACGCACAGACTCAGAGAASC GCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAG
GTGACGAAGAGCGTCTTGAGTTGTTATCACGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAATCACTTAGTCTCCTC

insert pLM1

ORF pLM1

T A S R R N S T I V L R T D S E K R S L A E S G L S V F S E S E E

AAAGCCCCATAAAAACTGGAGTACGACASTGGTAGCCTSAAGATGGAACCTTGGGACTTCTAAGTGGCGGAGGGAGCGGCC TGAAGCTGTGATGATTCAT
TTTCGGGGATTTTTTGACCTCATGCTGTACCAACCGGACTTCTACCTTGGACCTGAAGATTCACCGCTCCCTCGCCGGACTCTCGACACTACTAAGTA

insert pLM1

ORF pLM1

K A P K K L E Y C S G S L K M E P G T S K V R R E R P E S C D D S

CCAAAGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCTGGTTCCCTGAAGAAGGSCAAGACCCACCTGTGGCTGTAAC TCCCCATCACTCA
GGTCCCACTCTTGACTTTTTCGGGTAGTCGGACCCGCTGGGACCAAGGACTTCTTCCGTTCTGGGGTGGACACCGACATTGAAGGGGCTAGTGAGT

insert pLM1

ORF pLM1

G G E L K K P I S L G P G S L K K G K T P P V A V T S P I T H

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

CACAGCCCAGAGTGCCCTCAAAGTCGCASGCCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAACTACTGGGCTCCAAACGCTCCTCC
GTGTGGGTC TCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCGTGCCCATTCGAACGTCAC TTCTATGACCCGAGGTTGCGAGGAGG 240

insert pLM1

ORF pLM1

T A Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S

TCTGATGCTGGTGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTC
AGACTACGACCAAGCCCTGGCGGACTCACTACGATTCTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAG 250

insert pLM1

ORF pLM1

S D A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P

CTGCCACAGGCACAGCCACTGTTCATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGG
GACGGTGTCGCTGTCGGTGACAGTACGTTTGACCACCAAGTGGTGAGAGTCGTTCTAGGTC TTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGC 260

insert pLM1

ORF pLM1

P A T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCTTGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGSCCAGCCAA
GTCTGATCGAATCTACAAAGGTTGTCGCTCTCGGTCTTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTCGCTC 270

insert pLM1

ORF pLM1

K T S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A

TCAAGTTCTATGAGCGTGACCGSCGGCGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCTCAGCACCAAGCAGGGAGGCCCTTAGGC
AGTTCAAGATACTCGCACTGGCCGCCCGCCCACTGGAGCGGGACAC TCGTCGTCGTAAC TGGGGTCAGAGGAGTCGTGGTTCGTCCCTCGGAATGCG 280

insert pLM1

ORF pLM1

S S S H S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T

CTTCCAGACTGAAGGAGCCTACCAAGGTAGCCASTGSSCGACCACTCCAGCCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGC
GAAGGTC TGACTTCTCGGATGGTTCCATCGGTACCCGCCCTGGTGAGGTCGGGACAGTTAGTCTGTCTAGCCCTTTTCTCTTCCGCTTCGGTTCGG 290

insert pLM1

ORF pLM1

F S R L K E P T K V A S G R T T P A P V N Q T C R E K E K A A A

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

GAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCAATTGGTGGGCTGCCGTAATCCGATGACCAGTCAGAG
CTTCCCAGTCCATGGTCGAAGTCAGGGTCCCTCTCTGGTTCCCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTC

insert pLM1

ORF pLM1

K G L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D O Q S E

CTGCCCTTCCCCCTGCACCTCCCATGTCTCTGAGTGCAGGAGGCAACTTACCAACATAGTGAGTCCCACGCGGCCACACGCCAAGAATCACCCGCT
GACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCACGTTCCCGGTTGAATGGTGTATCACTCAGGGTGACGCCGGTGGTGCGGTTCTTAGTGGGCGA

insert pLM1

ORF pLM1

L P S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T P

CCAACAGCATCCCCACCCAGGCGGCCCTTCGAGCTGTACAGCGGCTCCCAATGGGAGCACCTGTCCCTGGCCGAGAGACCAAGGGAATGATTGG
GGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTCGCCGAGGGTTTACCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGC

insert pLM1

ORF pLM1

S N S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R

GTCAGGATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGTCTGCCCTCCAGTGCCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATG
CAGTCTTAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCTTAC

insert pLM1

ORF pLM1

S G S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M

CAATCTGAGCAAAATCCGGAAGCTTCGTAGGGAACGGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTCTGCCAATGCTAATCTGCTGG
GTTAGACTCGTTTAGGCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTACCCTGGGAAGTGCAGAGTCGAAGACGGTTACGATTAGACCAAC

insert pLM1

ORF pLM1

D S E Q I R K L R R E L E S S Q E K V A T L T S O L S A N A N L V

CTGCTTTTGGCAGAGGCTGGTGAATATGACATCCCGCTTGCAGACCTTGGCAGAGACGGCGAGGAGAAGGACACTGAGCTGCTGGATTGGGAGAAAC
GACGAAAACCTCGTCTCGAACCACTTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTCTCTGTGACTCGACGACCTAAACGCTCTTTG

insert pLM1

ORF pLM1

A A F E Q S L V N M T S R L R H L A E T A E E K O T E L L D L R E I

Tuesday, 18 November 1997 11:48
lig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTGAGGAGCCCTTAA TGCC TCAGAAACCACACCCAAAGAACTTCGGATCAAGAGA
GTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTCT 420
-----insert pLM1-----
-----ORF pLM1-----
I D F L K K K N S E A Q A V I O G A L N A S E T T P K E L R I K R
CAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAAGAAAAAAGA 430
GTTTGGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTCTTCTTTTCTTCT 440
-----insert pLM1-----
-----ORF pLM1-----
C N S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K
GTTGGGTCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTAC 450
CAACCCAGATACCGAAGCTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATG 460
-----insert pLM1-----
-----ORF pLM1-----
S V V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T
ACCCGACTCTTCAGCCCCCTCATCCCCAACTACAGCATGGTTCACAGAGACTGCTTCACCCCTCCATCAAGTCTCCACCTTGCTCTCGTGGGCACT 470
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGTCCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGA 480
-----insert pLM1-----
-----ORF pLM1-----
P D S S A P S S P K L C H G S T E T A S P S I K S S T L S S V G T
GATGTACCGAGGGCCCTGCTCAGCCAGCCCCACACTAGGCTGTTCCATGCAATGAGGAGGAGGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCT 490
CTACAGTGGCTCCCGGACGAGTGGGTCGGGGGGTGTATCCGACAAGGTACGTTTACTCTCTCTCGGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 500
-----insert pLM1-----
-----ORF pLM1-----
C V T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R
CTGAGCTATGGGAGAAGGAAATGAAGCTTACAGAGATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACAT 510
GACTCGATACCCCTCTTCTTTACTTTCGAATGCTCTGAGGCGAAGCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCTCTGGTACGTGTTGTA 520
-----insert pLM1-----
-----ORF pLM1-----
S E L W E K E M K L T C : R L E A L N S A H O L D O L R E T I H N F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

GCAGT TGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGSCAGGTCCCTGGATCATCT
CGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGSC TGACTTCCATCGGGTCCGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGACCTAGTAGA
420
-----insert pLM1-----
-----ORF pLM1-----
D L E V D L L K A E N D R L K V A P G P S S G S T P G Q V D G S S
GCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTACCCATTCTTCGGCCCCAGTCTGCAGACACAGACCTGTACCCATGGATGGCATCAGTA
CGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCCGGGTCAGAACGTCTGTGTC TGGACAGTGGGTACCTACCGTAGTCAT
480
-----insert pLM1-----
-----ORF pLM1-----
A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S
CTTGTGGTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAGGGGACTTGAAGCAGCAGGAATCTTCCTGGG
GAACACCAGGTTTCCTCTTCACTGGGAGGCCACCACTCTTACGGGGCGTCTGTAGTAGTTTCCCTGAACCTTCGTCTCTTAAAGAGGACCC
500
-----insert pLM1-----
-----ORF pLM1-----
T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G
CTGTAGCAAGGTCAAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCAAGTGTCAAGGACTATATTTCTAAATGGACCCAGCCCTCTACC
GACATCGTTCAGTCACCTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTTCAAGTTCTCTGATATAAAGATTTTACCTGGGTCGGAGATGG
510
-----insert pLM1-----
-----ORF pLM1-----
C S K V S G K V D V K H L D E A V F Q V F K D Y I S K M D P A S T
CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCCGAGATGGCTCCTTGGGTCGAGGTG
GAGCC TGAATGGTCACTCAGGTAGGTACCGA TGTCTAGTCGGTGCACCTTCTCACAACCTACGCTCTGGGGGGCTCTACGGAGGAACGGCAGCTCCAC
520
-----insert pLM1-----
-----ORF pLM1-----
L G L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G
TCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACACCTGGTGTTCAGAGCGCTGATCCCCAAGCCGATGATGCAGCACTACAT
AGTTATTTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGCAGCTGTGGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGTATGA
530
-----insert pLM1-----
-----ORF pLM1-----
V N N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

ATGCTTCCTGCTGAAGCACCGGCGCCTCGTCCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGCTGGAGCGC
TTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTGCGCGTGCCCGTTCTGGAAGGACGGTTAGCGAACCAGCTCATGGACCACCTCGC

insert pLM1

ORF pLM1

S L L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R

TCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACAGCAGTCTTGAAGGATCTGCAACTGTATCTTCCAACCTAGCCAACAGAG
AGACCGGCACTCCAGTGTC TCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCTGACAGACGTTCC TAGACGTTGACATAGAAAGGTTGGATCGGTTGGTC

insert pLM1

ORF pLM1

S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q

TAGACCGGGAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC
ATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCAC TAAGATAACCTACTGGACTCAC TTGCTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTG

insert pLM1

ORF pLM1

I D R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T

CTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC
GACGTTTCATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAACGTGAAC TCGAAGTCTTACAACCTG

insert pLM1

ORF pLM1

C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T

TTCTCCAACAACGTGGAGCCAGCCAATGGCTTCTTGGTTCGT TACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGC
ATGAGGTGTTGCACCTCGGTGGTTACCGAAGGACCAAGCAATGGACTCCTCTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTCTCTG

insert pLM1

ORF pLM1

F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E

TGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGTTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCT
ACCAAGCCCAAGAGCTGACCATGGGTTTCGACACCATAGTAGAGGTGTGGAAGGAACCTCTCGTGCTGTTGAGTCTGAAGGAGTACCGGGAACGAAGAA

insert pLM1

ORF pLM1

L L R V L D V V P K L V V H L H T F L E K H S T S D F L I G P C F F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1c

TCTGTGCTGTCCATTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACC TGTGGAACAAC TCTATCATCCCTATCTACAGGAAGGAGCC AAGGATGGG
 AGACAGCACAGGGTAACCGTAAC TCTGAAGGCTGGACCAAGTAAC TGGACACC TTGTTGAGATAGTAAGGGATAGATGCTCTCC TGGGTTCTCTACCC

- insert pLM1

-ORF pLM1

L S C P I G I E D F R T W F I D L W N N S I I P Y L Q E G A K D G

ATAAAGGTCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTG-
TATTCCAGGTACCTGTCTTTCGACGAACCTCTCTGGGTCACTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTGTTCTGGTTAGTTTCGACA

- insert pLM1

— ORF pLM1

I K V H G Q K A A V E D P V E V V R D T L P V P S A Q Q D Q S K L

ACCACCTGCCCCACCCACGGTGGGCCC TCACAGCAT TGCC TCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTC TCTGGAC TCAGATCC
 TGG TGGACGGGGGTGGGTGGCACCCGGGAGTTCGCTAACGGAGTGGAGGGC TCCTATCTGT CAGTTTCTGTCGTGGGGTCAAGAGACCTGAGTCTAGG

- insert pLM1

- OAF pLM1

Y H L P P P T V G P H S I A S P P E D R T V K O S T P S S L D S D F

TCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCTCGGACCCCAACCTTCAGGCAACACTTTAA
AGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATT

- insert pLM1

- ORF pLM1

L M A M L L K L O E A A N Y I E S P D R E T I L D P N L Q A T L

GGGTTGCGCAATCACTGTCACCCCGGACAGCAGAAAGCTGGGCATCAGCTATCTTAGCTCCTCCTCTCCCTCTCCTCTTTCAGAGCAC TGGCTCTCCAG
CCCAGCCGTTAGTGACAGTGGGGGCCCTGTCGTCTTGGCACCCTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCCGASAGGTC

- insert pLM1

G F G N H C H P R T A E R V H Q L S . L L L S P L L F O S T G S F

CCCCAGGAGGAGAACAGGAGGGAGGAGSAGA-5AASAGSGGCACGGTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGC
GGGGTCC TCC TCTTGCTCTCCCTCC TCC TCTACTTCTCTCCCTGTCCAAGAACCCAGACA TGGAAACTCTTGAAGGATCCCTTCTTACCACCCCACCS

- insert pLM1

A P G G E Q E G G G D E R S G T G S W C C T F E N F L G R N G G V A

G T T T G G G A A C T T G T G C C C C T A A A C A C A T T T A C T G G C C T C C T C T A A T G A C T T T G G G G A A A A G A T G A T T C T G G G T C T T C C C T T G A C T T C T T T T T C A A T T
 C A A A C C C T T G A A C A C G G G G G A T T T T G T A A A T G A C T G G A G A G A T T A C T G A A A C C C C T T T T C T A C T A A G A C C C A G A A A G G G A A C T G A A G A A C A A A G T T A A

- insert pLM1

F G N L C P L N T F T S L L . . L W G K O O S G S F P . L . V S I

Page 11

BNSDOCID: <WO 9824810A2 | >

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACCCAGCTGTGGAA
GTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTCCTTCTCAGGACTCCGCCCTTCTTGGTGGACACCTT
H S N M Y P L M R Q . P . . M L Q . Y . K R K S P E A E R T S C G 760

TGTGTGCAGTTAGGGTGTGGAAAGTCCCCAGGC TCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
ACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCTCCGCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGG
M C V S . G V E S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P 770

CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGC
GGTCCGAGGGGTCTGCTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGCGGGGTAGGGCGGGGATTGAGGGC
Q A P Q Q A E V C K A C I S I S Q Q P . S R P . L R P S R P . L R 780

CCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAAGGGCCGAGGCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGS
GGTCAAGCGGGTAAAGAGCGGGGTACCGACTGATTAATAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGCTTTCATCACTCC
P V P P I L R P M A D . F F L F M Q R P R P P R P L S Y S R S S E 790

AGGCTTTTTTGGAGGCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
TCCGAAAAACCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTCTACTCTAGCAAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGS
Kan/Neo 800

E A F L E A . A F A K I D Q E T G . G S F R M I E Q D G L H A G S P

GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCGTGTTCGGCTGTGACGCGCAGGGGCGCCCG
CCGGCGAACCCACCTCTCCGATAAGCCGATAGTACCCGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCGCGGGC
Kan/Neo 810

A A V V E R L F G Y D V A Q Q T I G C S D A A V F R L S A Q G R P

GTCTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAAGTGAAGACGAGGCGCGGCATCGTGGCTGGCCACGACGGCGTTCCTTGCAGC
CAAGAAAAACAGTTCTGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCGGATAGCACCGACCGGTGCTGCCCGCAAGGAACGGCTC
Kan/Neo 820

V L F V K T D L S G A L N E L O D E A A R L S V L A T T G V P C A

CTGTGCTCGACGTTGTCTAGTGAAGCAGGAGGAC TGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCCTGTCTCTACCTTGCTCCTGCCGAGAA
GACACGAGCTGCAACAGTGACTTCGCTCTCCCTGACCGACGATAACCCGCTTACGGGCCCCGCTCTAGAGGACAGTAGAGTGGAAACGAGGACGGCTCTT
Kan/Neo 830

A V L D V V T E A G R S V L L L G E V P G Q D L L S S H L A P A E

AGTATCCATCATGGCTGATGCAATTCGGGGGCTGATACGCTTGATCCGGCTACCTGCCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGT
TCATAGGTAGTACCGACTACGTTACGCCGCCGACGATGCGAAGTACGGCGATGGACGGGTAAAGCTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCA
Kan/Neo 840

V S I M A D A M R R L H T L D P A T C P F D H Q A K H R I E R A R

ACTCGGATGGAAAGCCGGTCTTGTGCTAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAC TGTTCGCCAGGCTCAAGGCGAGCATGC
TGAGGCTACCTTCGGCCGAGAACAGCTAGTCTCTAGAGCTGCTTCTGCTAGTCCCGGAGCGCGGTTCGGCTTGACAAGCGGTCCGAGTTCGGCTCGTACG
Kan/Neo 850

R M E A G L V C Q D D S L D E E H Q G L A P A E L F A R L K A S P

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 13

CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAAATGGCCGCTTTTC TGGATTATCATGACTGTGGCCGGCT
GGCTGCCGCTCC TAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGGGAAAAGACCTAAGTAGCTGACACCGGCCGA 960

Kan/Neo

P D G E O L V V T H G D A C L P N I M V E N G R F S G F I D C G R L

GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGCGGGCGAATGGGCTGACCGCTTCCTCGTCTTTACGG
CCCACACCGCTGGCGATAGTCTGTATCGCAACCGATGGGCATATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCA 970

Kan/Neo

G V A D R Y Q D I A L A T R D I A E E L G G E V A D R F L V L Y G

ATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC
TAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAAGTCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGCGG 980

Kan/Neo

I A A P D S Q R I A F Y R L L D E F F . A G L V G S K . P T K R R

CAACCTGCCATCAGGATTTTCGATTCCACCGCCGCTTCTATGAAAGGTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGC
GTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACCTTCAACCCGAAAGCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGG 990

P T C H H E I S I P P P P S M K G V A S E S F S G T P A G . S S S A

GGGGATCTCATGCTGGAGTCTTCGCCCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA
CCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCCTTCTTGGGCGCGATACTGCGGTATT 900

G I S C W S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .

AAAGACAGAAATAAACGCACGGTGTGGGTGCTTTGTTTCATAAACGCGGGGTTCCGGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCATTS
TTTCTGTCTTATTTTTCGTCGCCACAACCCAGCAAAAGTATTTGCGCCCCAAGCCAGGGTCCCAGCCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC 910

K D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I

GGGCCAATACGCCCGGCTTTCTTCTTTTCCCCACCCACCCCCCAAGTTCCGGGTGAAGGCCAGGGCTCGCAGCCACGTCGGGGCGGCAGGCCCTGGC
CCCGGTTATCGGGGCGCAAGAAGGAAAAGGGGTGGGTGGGGGTTCAAGCCCACTTCCGGGTCCGAGCGTCGGTTGCAGCCCCGCCGTCGGGACGS 920

G A N T P A F L P F P H P T P Q V R V K A Q G S Q P T S G R Q A L F

ATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCA
TATCGGAGTCAATGASTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAAATATTAGAGTACTGGT 930

P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . F

AAATCCCTTAACGTGASTTTTCGTTCCACTGAGGCTCAGACCCGCTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTCTTCGCGTAATCTGCTG
TTTAGGGAATTCGACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTAGAGAAGTCTAGGAAAAAGACGGCATTAGAGCAC 940

pUC ori

K S L N V S F R S T E R Q T P . K R S K D L L E I L F F C A . S A

CTTGCAAAACAAAAAACCCGCTACCAAGCGGTGGTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAC
GAACGTTTGTTTTTTGGTGGCGATGGTGCACCAAAACAAACGGCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCAAGTCTCTCGCGTC 950

pUC ori

A C K Q K N H R Y Q R V F V C R I K S Y Q L F F R R . L A S A E R F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 14

ATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACTCGCTCTGCTAATCCTGTTACCA3
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCCTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTC 960

pUC ori

Y Q I L S F . C S R S . A T T S R T L . H R L H T S L C . S C Y Q

TGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGC TGAACGGGGGGTTCGT3
ACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTGGCAGCCGACATTGCCCCCAAGCAC 970

pUC ori

V L L P V A I S R V L P G V T Q D D S Y R I R R S G R A E R G V R

CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGCGGGAC 980
GTGTGTCGGGTGCAACCTCGCTTGTGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCTCTTTCGCGCT3

pUC ori

A H S P A V S E R P T P N . D T Y S V S Y E K A P R F P K G E R R T

AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTTTCGCCACC 990
TCCATAGGCCATTTCGCGTCCGAGCCTTGCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG

pUC ori

G I R . A A G S E Q E S A R G S F Q G E T P G I F I V L S G F A T

TCGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGGGCCTTTTACGGTTCTTGCCCTTTTGCT3
AGACTGAATCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGGTCGTTCGCCGGAAAAATGCCAAGGACCGGAAAAACGAC 1000

pUC ori

S D L S V D F C D A R Q G G G A Y G K T P A T R P F Y G S V P F A

GCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 10070
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
G L L L T C S F L R Y P L I L W I T V L P P C I

SEQUENCE ID No. 8.

Tuesday, 18 November 1997 10:34

fig 30 pEGFP72 (1>9697) Site and Sequence

Enzymes : 72 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Page 1

16p

100 TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTAAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

200 CCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTACGGT
GGGTTCGTTGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

300 AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCTAT
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

400 CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCCGTAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCAAAACCGTCATGTAGTTACCCGACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V

500 TAGCGGTTGACTCACGGGATTTCCAAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAATGTCGTA
ATCGGCAAACTGAGTGCCCTAAAGGTTCAAGGTTGAGGTAACGTCAGTTACCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTACAGCAT
! A V . L T S I S K S P P H . R Q V E F V L A P K S T G L S K M S .

600 ACAACTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAACGCTTTACCGCCATCCGACATGCCACCTCCAGATATATTGCTCTCGACCAATCACTTGGCAGTCTAGGCGATCGCGAT
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L

700 CCGGTCGCCACCATGTTGACCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
GGGACGCGGTGGTACCACTGTTCCCGCTCCCTGACAAAGTGGGCCCCACACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGCTGTTCAAGTCGC
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

800 TGTCCGGCGAGGGCGAGGATGCCACCTACGCAAGCTGACCTGAAGTTCATCTGCACACCGGCAAGCTGCCCGTGCCCTGSCCCACCTCGTGAC
ACAGGCGGCTCCCGCTCCCGTACGGTGGATGCGCTTCCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCCAGCGGCGACGGGACCGGTTGGGAGCACTG
V S G E G E S D A T Y S < L T L K F I C T T G K L P V P V P T L V T

eGFP.C.e.unc53

eGFP.C.e.unc53

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1

CACCC TGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGGACTTC TTCAAGTCCGCCATGCCGGAAGGCTAGGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCGGATGCAGGTCTCT 900

eGFP.C.e.unc53

T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

KspI

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACCTGGCGTAGCTCGACTTCCCGTAGC 1000

eGFP.C.e.unc53

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAGCCACAACGCTTATATCATGGCCGACAAAGCAGAAGAACGGCATCAA
TGAAGTTCTCTCGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTCTTTCGCCGTAGTT 1100

eGFP.C.e.unc53

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I A

GGTGAACITCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCTGCTTGTGGGGGTAGCCGC TGCCGGGACACGACGAC 1200

eGFP.C.e.unc53

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACAATGGTCTGCTGGAGTTCTGACCCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGTCAAGCGGGAAGCTGTTCTGGGGTTGCTCTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCAGGCGCT 1300

eGFP.C.e.unc53

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

BspM II

Bgl II

Asu II
EcoN I

TCACTCTCGGCATGGACGASCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCCTAACCCGGTTAGCGGTGGAAG 1400

eGFP.C.e.unc53

C.e. unc53

T L G N D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

Nru I

EcoR I

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAAATGTSATCGTTCCSATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGTTACTAAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACACTAGCAAGGCTASTTGCTT 1500

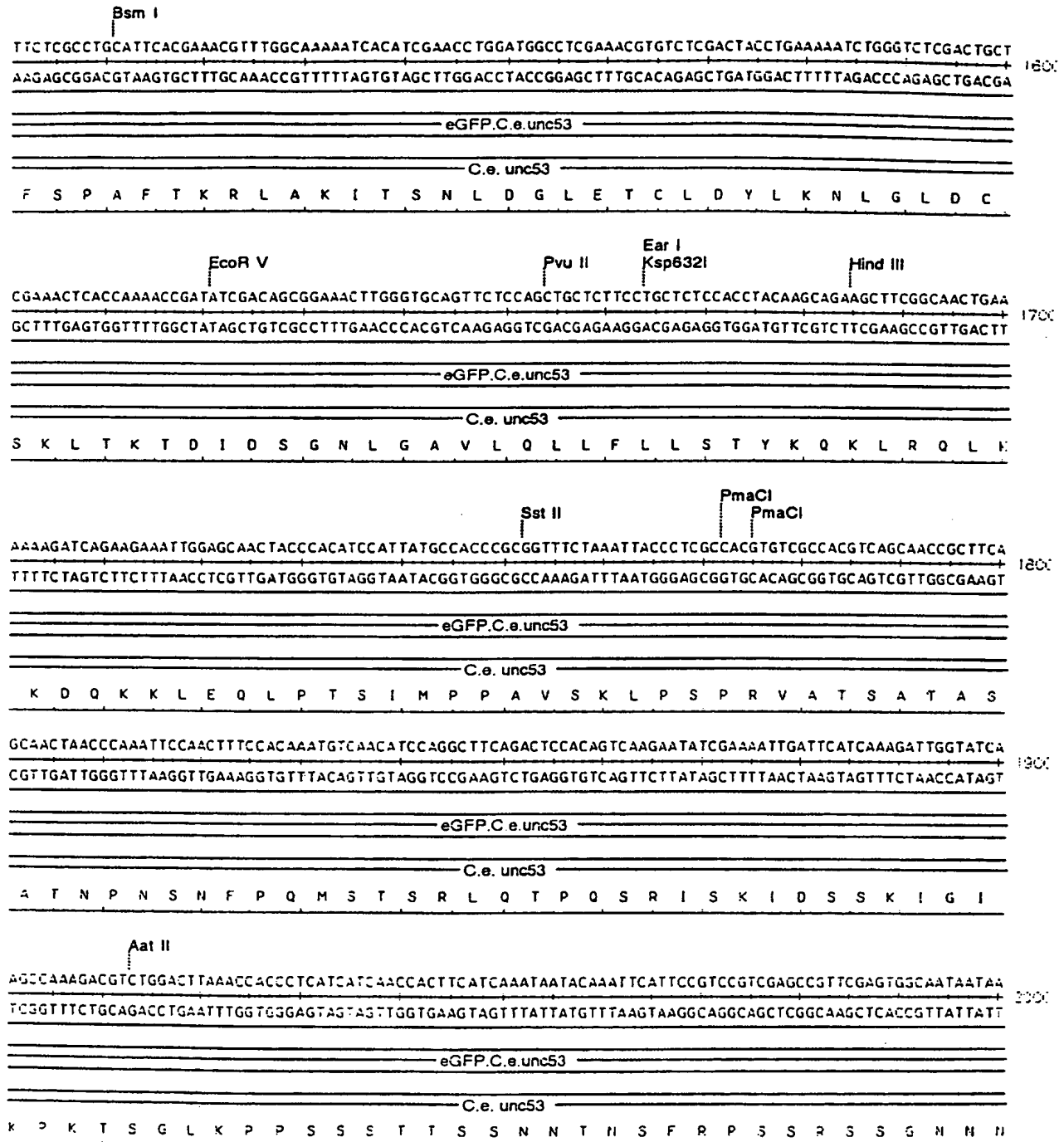
eGFP.C.e.unc53

C.e. unc53

T G S L S K A I P D I S N D F R D Y P L V S Q L I H A I V I N E

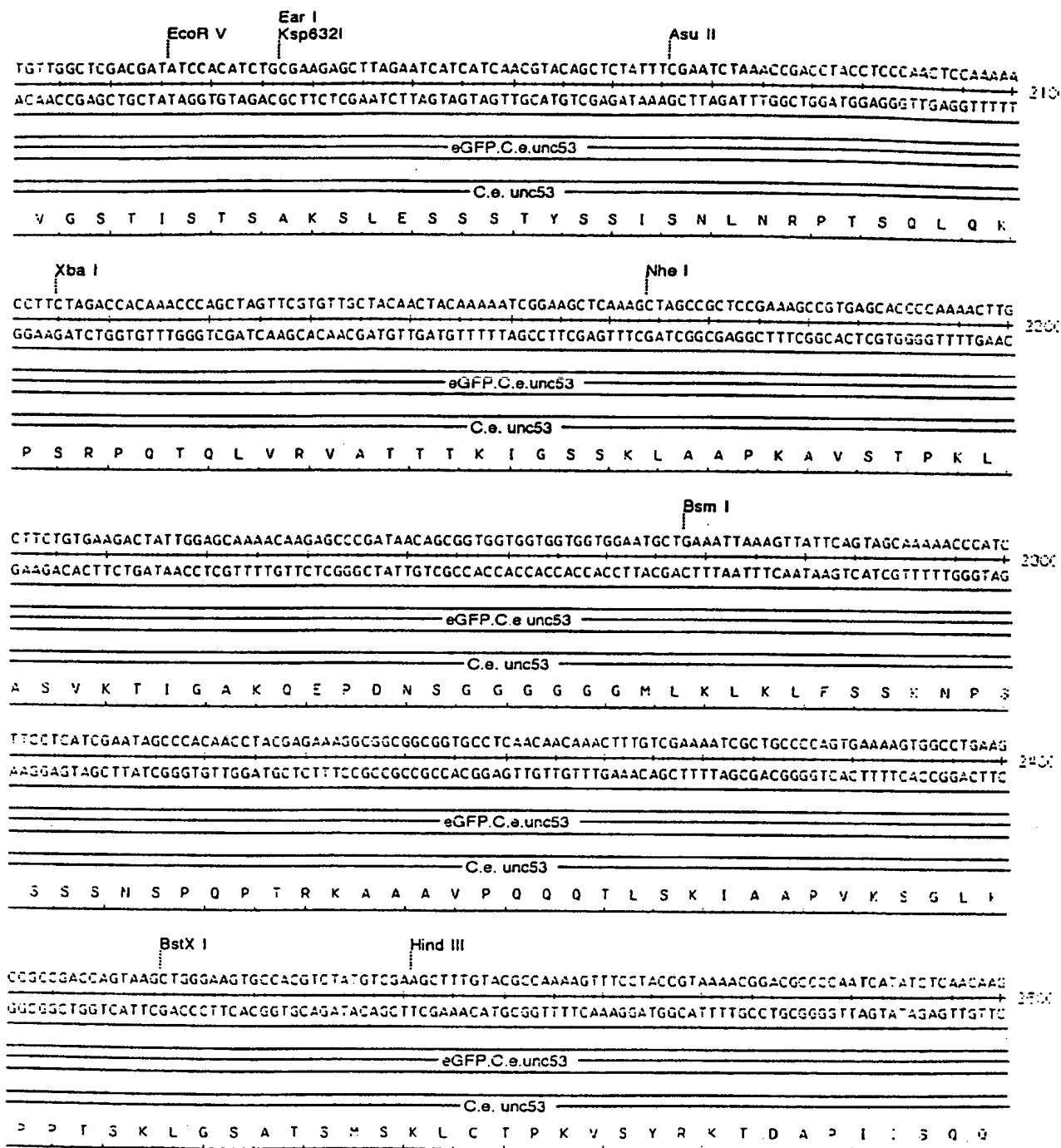
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 3



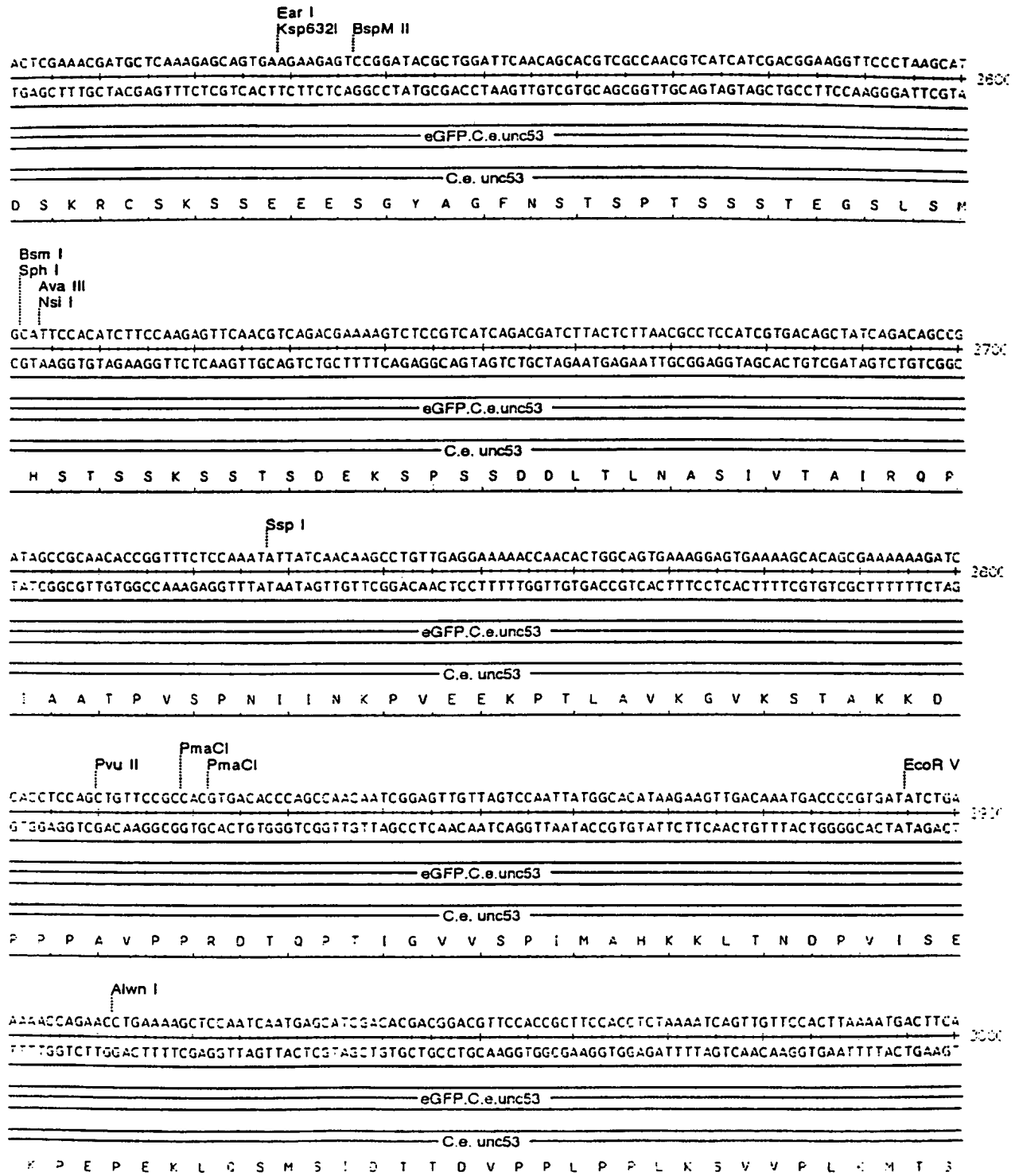
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4



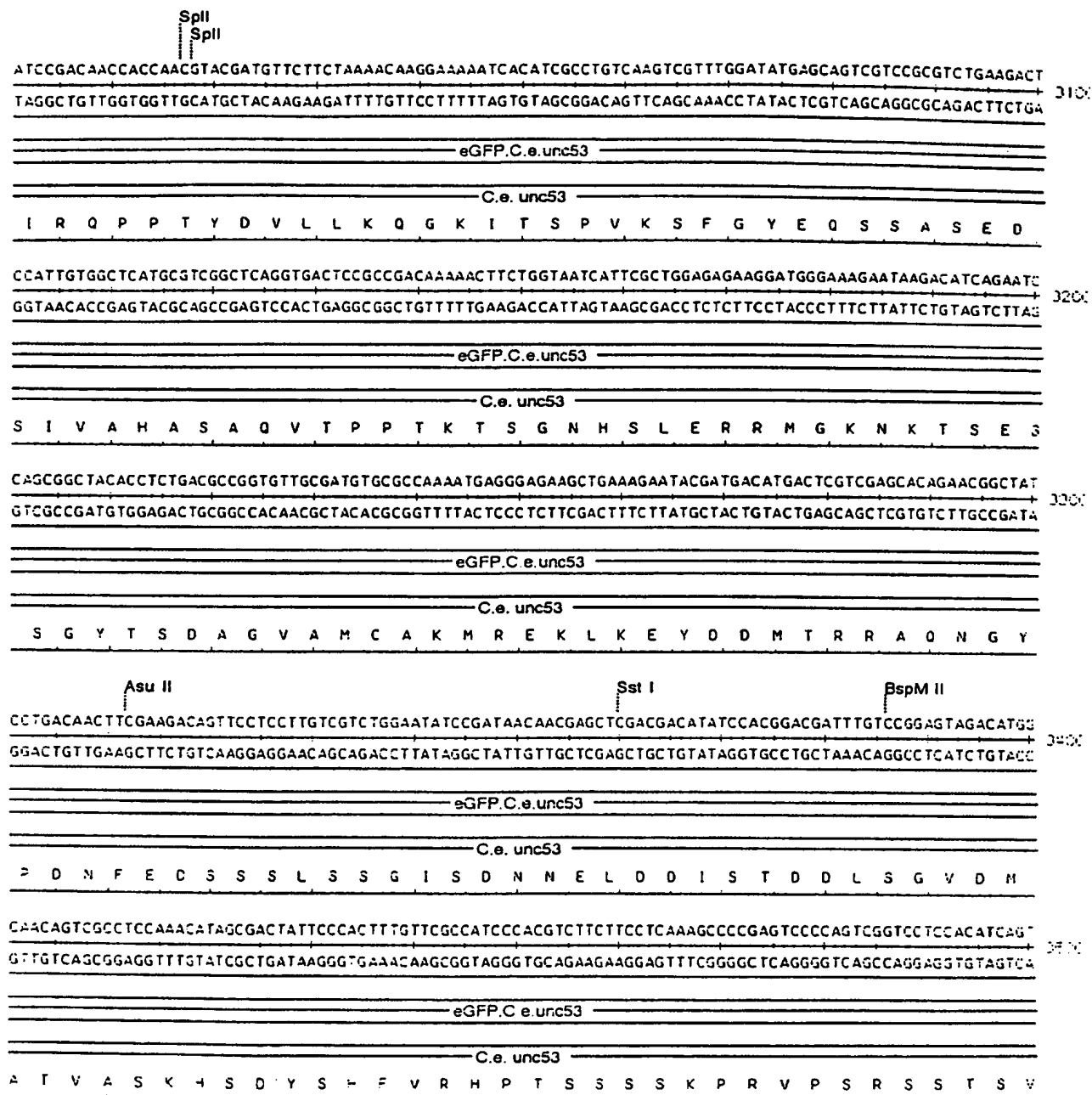
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 5



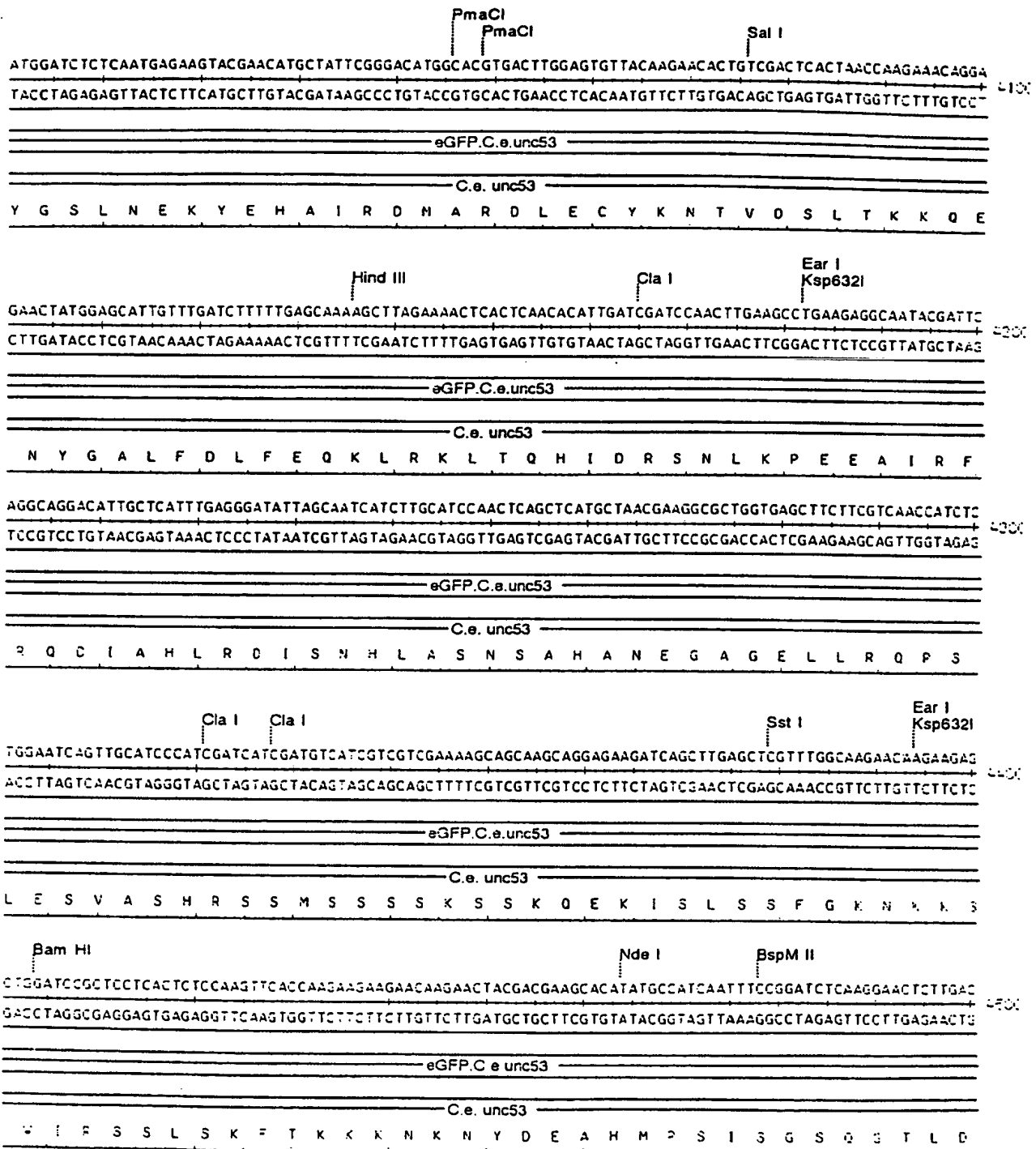
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 6



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 8



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4

Sst I Apa I
AACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACCTTACGAAGTCCGCCTTGACAATCTGGATCTGCCCGCGAAGTTGATG 4600
TTGTAAC TACACTAACTCAACTTCGTTCTCGAGTTTCTTGGCCTATCACGTGAAATGCTTCAGGCGGAACCTGTTAGACCTAGCACGGGCGCTTCAACTAC
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
N I D V I E L K Q E L K E R D S A L Y E V R L D N L D R A R E V D
TTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCACGGTCCAGCCACTCGTGCTTCTTCCCG 4700
AAGACTCCCTCTGTCACTTGTCAACTTTTGGCTCTTGTTCGTTAATTTCTTTCTTACCTGTTTGAGTGGTTGCCAGGTCGGTGAGCACGAGAAGGGC
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
V L R E T V N K L K T E N K Q L K K E V D K L T N G P A T R A S S R
Ksp I Bsr I Asu II
CGCCTCAATTCAGTTATCTACGACGATGAGCATGCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAAC 4800
GCGGAGTTAAGGTCAATAGATGCTGCTACTCGTACAGATACTACGTCGCACATCGTTCATGTAGTCGATCAGTTAGAAGCTTTGCTAGGAGACCGACGTTG
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
A S I P V I Y D D E H V Y D A A C S S T S A S Q S S K R S S G C H
Pvu I Hpa I EcoR V
TCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTTCGATCGTTAACC CGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCA 4900
AGTTAGTTTCAATGACATTTGCACCTGTAGCGACCTCTTTAGTCAAGCTAGCAATTGGGCCTGTTTCTCTATTAGCATCTATAGAAGGTAACAGTTGGT
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S I K V T V N V D I A G E I S S I V N P D K E I I V G Y L A N S T
Cla I
GTCAGTCATGCTGSAAGACATTGATGTTTCTATTCTAGSACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTTGGAAATCGATGCTCG 5000
CAGTCAGTACGACCTTTCTGTAAC TACAAGATAAGATCTGTATAAATTCAGATGGATAGGTCCTTAAC TACACCTCGTAGTTGAACCTTAGCTACGAGC
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S D S C W K D I D V S I L G L F E V Y L S R I D V E H Q L G I D A F

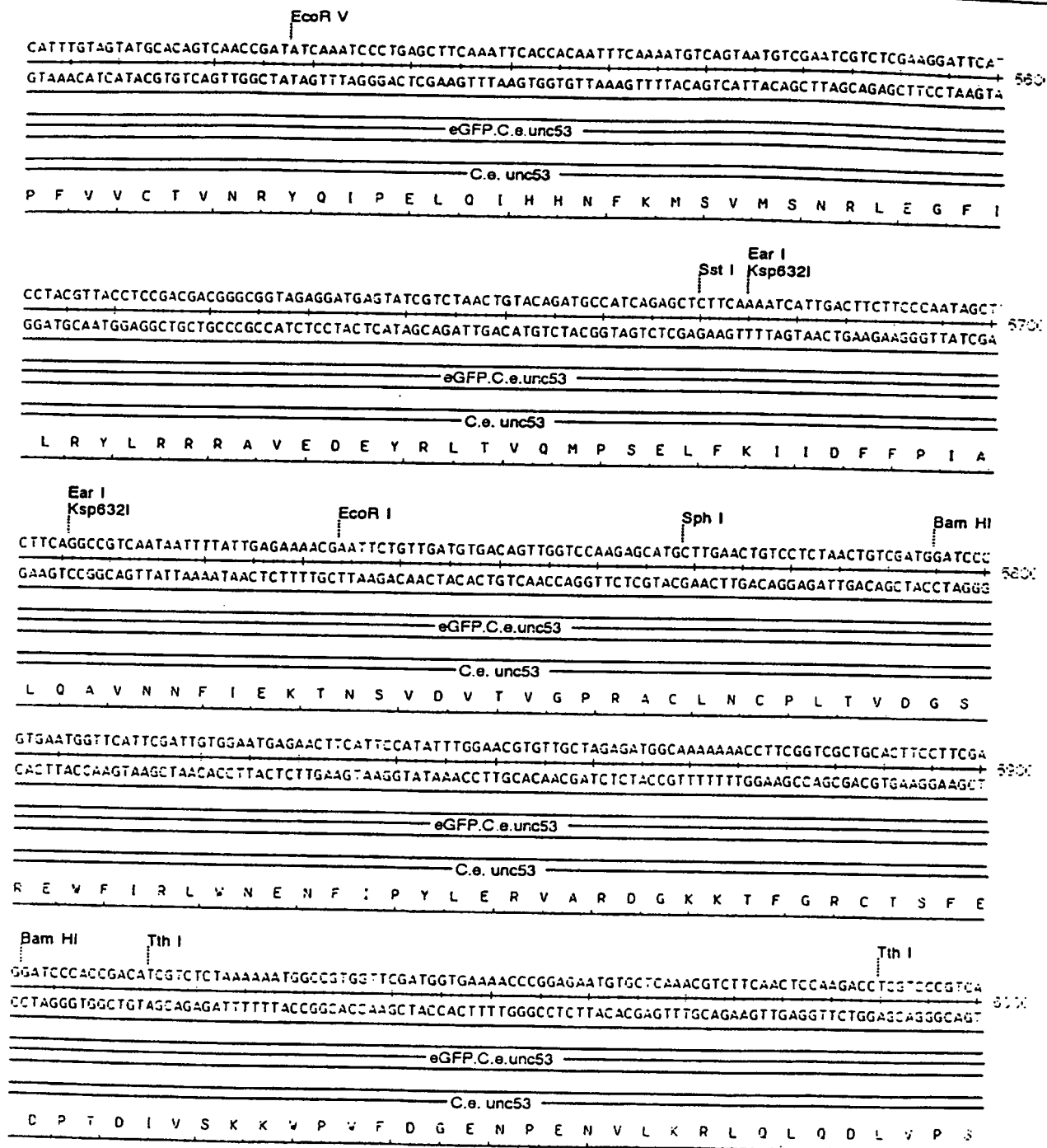
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 10

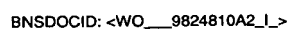
Mlu I
TGATTCTATCCTTGGCTATCAAATTGGTGAACCTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAACCTGACATTCTTACTTCCTCA
ACTAAGATAGGAACCGATAGTTTAACCACTTGAAGCTGCGCAGTAACCTCTGAGGTGTTGGTACTATTGGTCGGTAGGTTGACTGTAAGAATGAAGGAG 5100
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
D S I L G Y O I G E L R R V I G D S T T M I T S H P T D I L T S S
ACTACAATCCGAATGTTTCATGCACGGTGCCGCACAGAGTCGCGTAGACAGTCTGGTCTTGTATATGCTTCTTCCAAAGCAAATGATTCTCCAACTCGTCA
TGATGTTAGGCTTACAAGTACGTGCCACGGCGTGCTCAGCGCATCTGTCAGACCAGGAACATACGAAGAAGGTTTCGTTTACTAAGAGGTTGAGCAG 5200
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
T T I R M F M H G A A Q S R V D S L V L D M L L P K Q M I L Q L V
Aat II Bsr I Bsr I Asu II
AGTCAATTTTGACAGAGAGACGCTCTGGTGTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATGTATCTATTTCGAAC
TCAGTTAAACCTGTCTCTCTGCAGACCACAATCGACCTCGTTGACCTTAACCTTTCTCGTTTGACCGCTTCTGGGACCGACGAATACATAGATAAGCTTS 5300
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
K S I L T E R R L V L A G A T G I G K S K L A K T L A A Y V S I R T
Bsm I Xmn I Bgl II
AAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAAGAAGAATTGCTTCAAGTGGAAACGACGCTGGAAAAGATCTTGAGAGGC
TTTAGTTAGGCTTCTATCATAACAATTATAGTCGTAAGGACTTTTGTATTCTTCTTAACGAAGTTCACCTTGC TCGGGAACCTTTCTAGAACTCTTCG 5400
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
N Q S E D S I V N I S I P E N N K E E L L Q V E R R L E K I L R S
Ava III Nsi I Xba I
AAAGAATCATGCATCGTAATCTAGATAATATCCCAAAASAAATCGAATTGCATTGTTGTATCCGTTTTTGGAAATGTCCACTTCAAAACAACGAAGGTC
TTCTTAGTACGTAGCATTAAAGATCTATTATAGGGTTTCTTAGCTTAACGTAAACAACATAGGCAAAAACSTTTACAGGGTGAAGTTTTGTTCCTCCAG 5500
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
K E S C I V I L C H I P K N R I A F V V S V F A N V P L Q N N E G

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 11

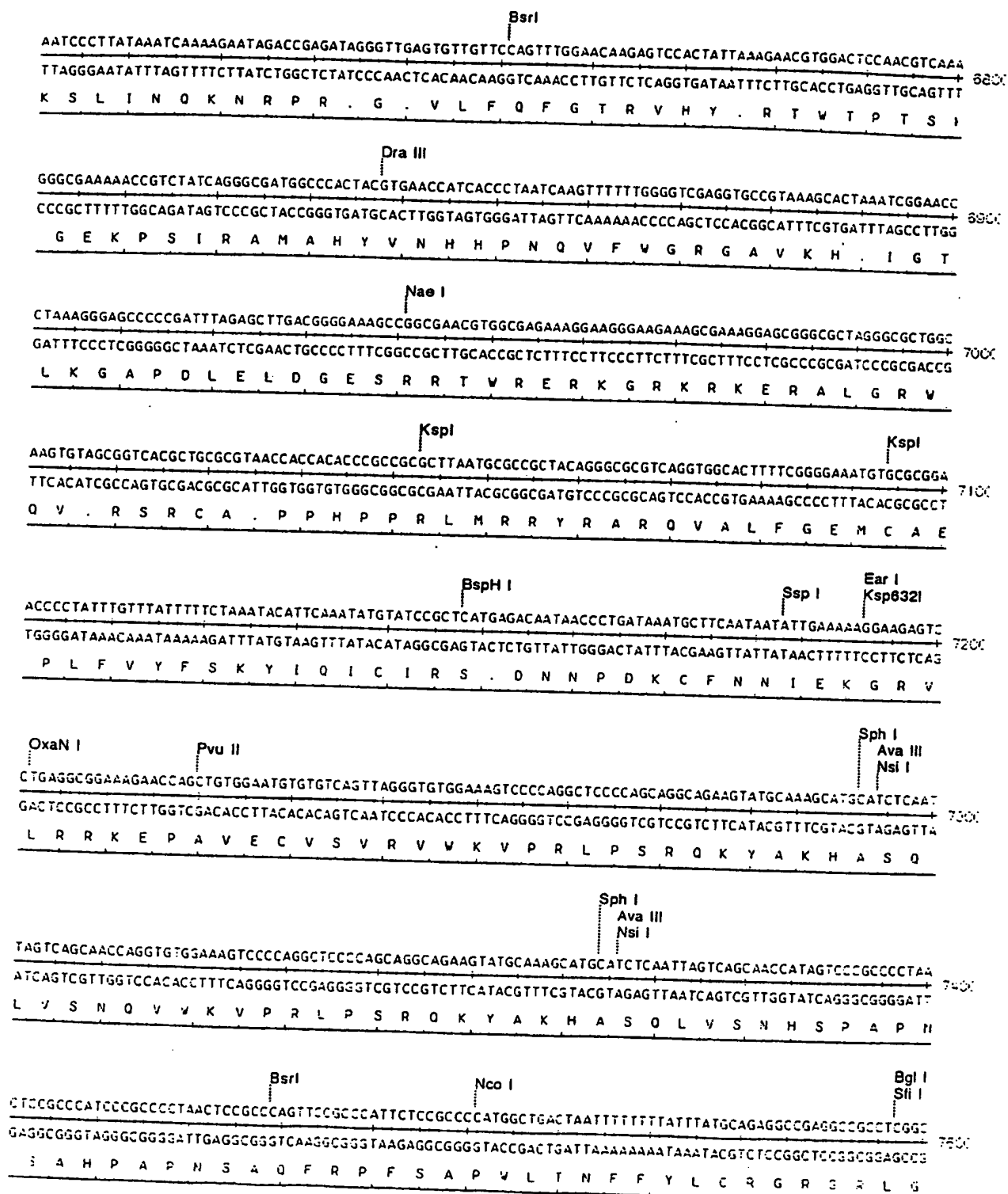


Page 12



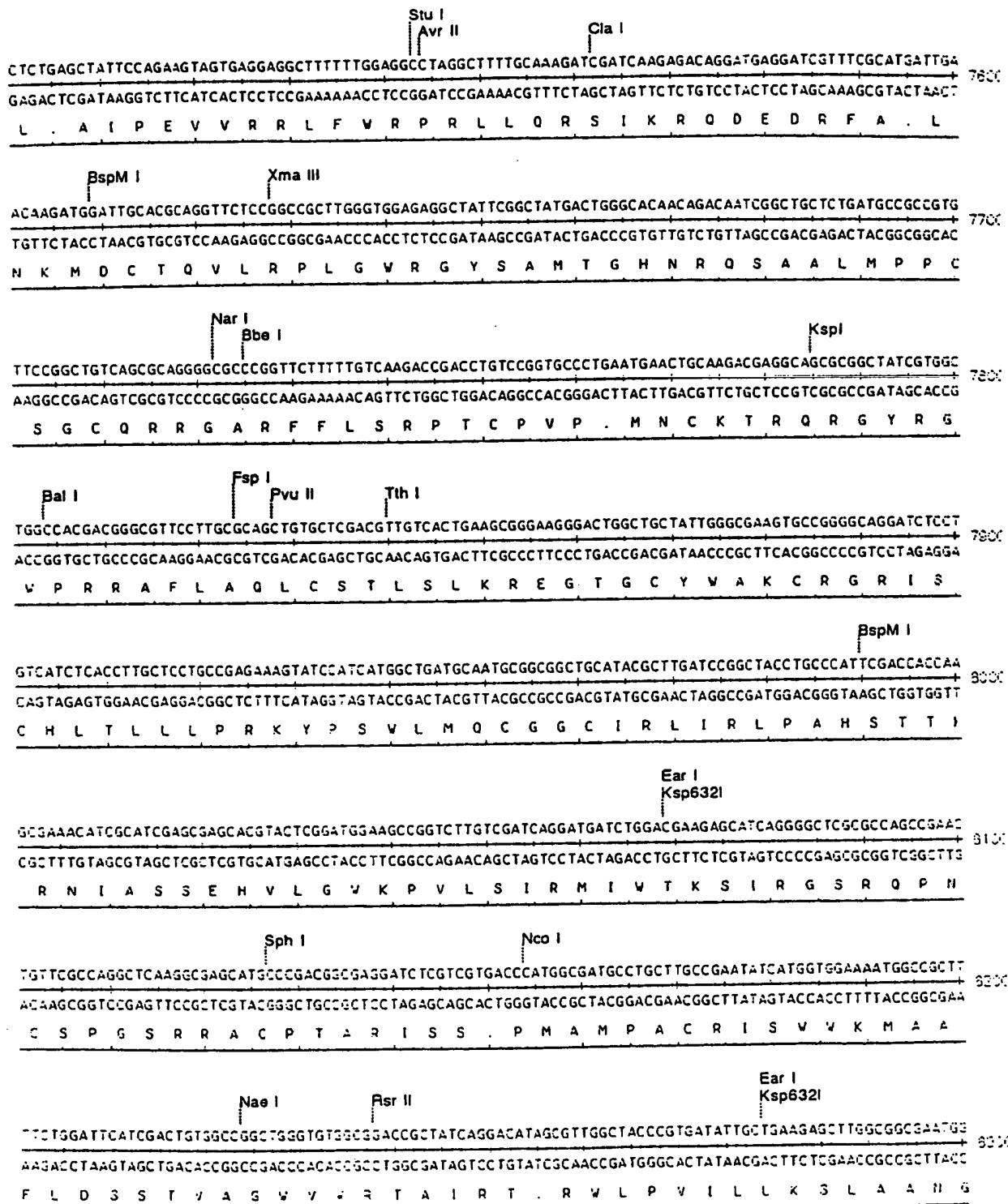
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 13



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 14

GCTGACCGCTTCCTCGTGTTCACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCATCGCCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGG
CGACTGGCGAAGGAGCAGCAAATGCCATAGCGGCGAGGGCTAAGCGTTCGGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCC
L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G 840

Asu II BspM I
GTTGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGAGATTCGATTCCACCGCCGCTTCATGAAAGGTTGGGCTTCGGAATCGTTTCCGG
CAAGCTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCAACCCGAAGCCTTAGCAAAAGGCC
V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F P 850

Nae I KspI Avr II
GACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGG
CTGCGGCGGACCTACTAGGAGGTCGCGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCGATTGACTTTGTGCTTCTCTGTATGGCC
G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G 860

KspI
AAGGAACCCGCTATGACGGCAATAAAAAGACAGAATAAACGACGCGGTGTTGGGTGTTTGTTCATAAACCGGGGTTGGTCCCAGGGCTGGCACTC
TTCTTGGGCGGATACTGCGGTATTTTCTGTCTTATTTGCGTGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCAGCGTGAG
R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L 870

TGTCGATACCCACCGAGACCCATTGGGGCCAATACGCCCGCTTCTTCCTTTTCCCCACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCA
ACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATSCGGGCGCAAGAAGGAAAAGGGGTGGGGGTGGGGGTTCAAGCCCACTTCGGGTCCCGAGCGT
C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A R 880

AlwI OxaI Dra I Dra I
GCCAAGCTCGGGGCGGAGGCCCTGCCATAGCCCTCAGGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGA
CGGTTGACGCCCCGCGCTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAATTTTCTAGATCCACT
S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E 890

BspH I
AGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTCTTGAGA
TCTAGGAAAACTATTAGAGTACTGGTTTTAGGGAATTGCAC TCAAAAGCAAGGTGACTCGCAGTC TGGGGCATCTTTCTAGTTTCTAGAGAAGACTCT
D P F . . S H D O N P L T . V F V P L S V R P R R K D Q R I F L R 900

TCTTTTTTCTGCGCGTAATCTGCTGCTTGCAACAAAAAACCCCGCTACCAGCGGTGGTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCG
AGGAAAAAAGACGCGCATTAGACGACGAACGTTGTTTTTGGTGCGGATGGTCCCAACCAACCGGCTAGTTCGATGGTTGAGAAAAAGGC
S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S 910

BsrI
AAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAACTACTGCTCTTAGTGAGCGTAGTTAGGCCACCACTTCAAGAAGCTCTGTAGCACCGCTACAT
TTCCATTGACCGAAGTCGCTCGCTCTATGGTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTA
E G N W L Q Q S A O T < Y C P S S V A V V R P P L Q E L C S T A Y I 920

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 16

AlwI
ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA 9300
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTACGACAGAAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGT
P R S A N P V T S G C C Q W R . V V S Y R V G L K T I V T G . G A

Apal I
CGGGTCGGGCTGAACGGGGGGTTCTGTCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC 9400
CGCCAGCCCGACTTGCCCCCAAGCAGGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTGCGACTCGATACTCTTTCGCGG
A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K R

ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCTGGTATC 9500
TGGGAAGGGCTTCCCTCTTTCGCGCTGTCCATAGGCCATTGCGCGTCCAGCCTTGCTCCTCTCGCGTGC TCCCTCGAAGGTCCCCCTTTCGCGGACCATA
H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V S

TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC 9600
AAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTGCGGTCTGTGCGCCG
L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R G

Ava III
Nsi I
CTTTTACGGTTCCCTGGCCTTTTCTGGCCTTTTCTCACATGTTCTTTCTCGGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 9697
GAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGACAAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Monday, 1 December 1997 14:12
ig PCB501

SEQUENCE I.D. No 9

Page 1

10 20 30 40 50 60 70
ATGACCAIGATACGCCAAGCTTGCAAGCTGAGGAATTGATATCAAGCTTATCGATACCGTCGACCT 70
GAGGATCAGAAAGAAATTCGACCAACTACCCACATCCATTAAGCCACCCGCGGTTCTAAGTGAGTTAA 140
TTTGGAGTTACGACACAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGCTATTCTGTATGACT 210
AGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAGGAACAATATTTACTTTGCTTATTTCGCTAAC 280
TTTGATTTAGTTTTCGATCAACTAGATCTTACAAAACCTTCAATACAATTCCATTTTCAGATTACCTC 350
360 370 380 390 400 410 420
CCACGCTGTCGCCACGTCAGCAACCGCTTCAGCAACCAACCCCAAATTCACAACTTTCACAAAAIGTCAACA 420
TCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTCGTAAGAAATTTTATTTTGGAGTCAAACTTGT 490
ATAAAATGCCAGAAAAGAAGATGATAAAAATGTAGTTTTCGCAAAAACCTCCACCTTTATTGCTCTAA 560
TATGACGGCTTATATCTCAATTTCTTASATTTTATCAAAAAATTTTCACATATACAAAATGTAAGAAAG 630
ATTTTCACAAAAATTTTTCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAAGCTGTTAA 700
710 720 730 740 750 760 770
AGTGGAGGAGCGCAAGTCTATACATGCAAAATATGATCTGAAACAAATTTGTGCTATTCTCAAAATGTTTA 770
AGCATGTTTTGAAGATTTTTCAAAATTCGCACTAGTTTCAGAACCTTCCTTTTGTATGAAAAAGTAAA 840
AAAAAACTATTTCAAAACCTCACCGCCACCATGTTTCACACTCTTAATTTTATAAAATTTTGCATTTAC 910
AAATCGCCCTCCCTTGGCGGAAAAGTCCCGACCAAAATCAATTTCTCGGCTTCATAATGACATTTAAAT 980
CATGTCAGAAAAACACAGAAAGAGGCTAACTAAATTCACAGGGACAGGTTCGCTCTCTTCGCTCTCTCT 1050
1060 1070 1080 1090 1100 1110 1120
CCGCTTCCTCTCTCTCTTCCATCTCCAACAACAACAATTTCCAAATTTGCTTGTCCATTTTCGCTTAA 1120
CATTTGTTGTTGGAAGGAACTACACCGGAGAGCGGTCAATTAATTCGAATGAGAGCAATGGCAATTCATC 1190
TTTCGGAAATTTGATGAATAAAGATACACCGGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGCTT 1260
TTTCATCGCTCTCAACTTGGGCAAGAGTTTCGCGCGCTCTCATCACTGACAATTAATGTGCGGTTTTATG 1330
CGCTCTTTCTCTATTCGCGCACTCATCTGAGTTACCAAACTGGAATACATTTTACACTATTCGAAGCC 1400
1410 1420 1430 1440 1450 1460 1470
ATTTATTTTCATATTTAAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAGTTTATTTAAAAAAGC 1470
ATATTTTCGATTTTAAAAATTCGAAAGTTTCAAAAAATCAATTAATTTTCCCAAAATTTGTATGCC 1540
TAAATTTTATTTCTACTGTTGACAAATATCTTTATATATGATCACTGTTTTCCATCTCAAAACCTTGAA 1610
CGCAAGTTATAGGAACCTCCGTGTCAATTTCCCATGCTATGAAATCGCTACTCAGCACATATCCAAAA 1680
TTAAGCTAGACGGTTGATAATTAATTCGCAAGCGGTAATAAAGTCAAGCAGTTAGAATTTTAATTCAAGC 1750
1760 1770 1780 1790 1800 1810 1820
ACAGATTATCTATCAAAATTCAAATTTTCAACATTCACCCAGTTCTGACAAATTTCCATGCTTTTTGGCC 1820
ATTAAAAAATTTTCACCTCTTCATCCATCTCACTCTGATCATAAAAAGTATAGCAAAAGCCGACTCT 1890
ACTTTTTAAGAGAAGGAGATCTGAGGCACATGGCGTGTGACCTTTTCATCTCGTCCGTTGGGCTCAA 1960
ATTACCGCTCAATCACTCTTCATATAGCATAGACCCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 2030
TATTTTCTGCTGAAAGCCGGGAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100
2110 2120 2130 2140 2150 2160 2170
AAAAACGAGSCAAATTAATAATATTAAATTAATGAGTTTGTAGATGAGATTTGGAAAAGAAGAAAA 2170
ATAAAAAATAGGAACCGCCAGATCAAAATTTCTATTTAAAGGTTTCAAGATCTTAGGCAAGATTCCG 2240
CTAACACAAAACTGAAGTGGCTGCAATTAATCTAGTGTAACTTTAGATTGAAGTGGGAAATCTAAGCC 2310
TGAATATAGCTTATCTAGATCTTGTGCAATCAAGCCCAAGCAGAAATGACATTCATTTA 2380
GTAAAGCCTAGATGACTTCT 2450

Page 2

BNSDOCID: <WO__9824810A2_I_>

Monday, 1 December 1997 14:12
fig pCB501

Page 3

4910 4920 4930 4940 4950 4960 4970
A"TCAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAAC 4970
GAAGGCGCTGGTGAAGCTTCTTCGTCAACCACTCTGCAATCAGTTGCATCCCATCGATCATCGATGTCAT 5070
CGTCGTGAAAAGCAGCAAGCAGGACAAGATCAGCTTGAGCTCGTCTGGCAAGAACAAAGAGAGCTGGAT 5170
CGGCTCCTCACTCTCCAAGTTCACCAAGAAAGAAACAAGAACTACGACGAAGCACATATGCCATCAATT 5180
CCGGATCTCAAGGAACCTCTGACAACTTGTGATGATTGAGTTGAAGCAAGAGCTCAAGAAAGCGGATA 5250
6280 6270 6280 6290 6300 6310 6320
GTGCACTTTACGAAGTCCGCTCTGACAATCTGGATCGTSCCCGCGAAGTTGATGCTCTGAGGGAGACAGT 5320
GAAGAAAGTCTGAAAALLGAGAAALAGCAATTAAGAAAGAAAGTGGACAAACTCAACCAAGGCTCAGGCACT 5390
CGTCTCTTCTCCCGGCTCAATTCAGTTATCTACGACGATGAGCATGTCTATGATGTCAGGCTGTAGCA 5460
GTACATCAGCTAGTCAATCTTGGAAAGCACTCTGGCTGCAACTCAATCAACGTTACTGTAAACGTTGGA 5530
CATCGCTGGAGAAATCAGTTGATCGTTAAGCGGACTTGAAGCAGCAGGAATCTTCTGGGCTGTAGC 5600
5610 5620 5630 5640 5650 5660 5670
AAGGTCAGTGGAAAAGTCTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACATATCT 5670
CTAAAATGGACCCAGCTCTACCTCCGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGT 5740
GAAACGAGTGTGGATGACAGACCCCCCGAGATGCCCTCTGCGCTGAGGCTGCAATTAACATATCAGTC 5810
TCCCTCAAAGGTCTGAAAGGAGAAATGCTGACAGGCTGCTGTTGAGAGCGCTGATCCCCAAGCCGATGA 5880
TGCAGCACTACAATAGCTCTCTGCTGAGACCGGCGCTCTCTCTCGGCCCCAGCGGCAAGGGGCAA 5950
5960 5970 5980 5990 6000 6010 6020
GACCTACCTGACCAATCGCTTGGCCGACTACCCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATC 6020
GTCAGCACTCTCAACATGCAACAGCAGTCTTGCAGGATCTGCAACTGTATCTCTTCAACCTAGCCCAAC 6090
AGATAGACCGGAAACAGGAATGGGGAATGCTCCCTGCTGATCTATTGGATGACCCTGAGTGAAGCAGG 6160
CTCCATCAGTGAAGTTGCTCAATGGGCCCCCTCACTGCAAGTATCATAAATGTCCCTATATTAAGGTAAC 6230
ACCAATCAGGCTGIAAAAATGACACCCCAACCATGGCTGCACTGAGCTTCAGGATGTTGACCTTCTCCA 6300
6310 6320 6330 6340 6350 6360 6370
ACAACGTGAGGCCAGCCAATGGCTCTCTGCTCTGCTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCA 6370
CATCAATGCCAACAAAGGAAGAGCTGCTCTGGGCTGCTGAGCTGGGTACCCCAAGCTGTGGTATCATCTCCAC 6440
AGCTTCTCTGAG 6510
GCATGAGGAGCTTCCGGACCTGGTTCATGAGCTGTGGAACAACCTCATCATCTCTTCACTACAGGAAGG 6580
AGCCAAAGGATGGGATAAAGGCTCAATGAGACAGAAAGCTGCTTGGGAGGACCCAGTGAATGGGTCCGGGAC 6650
6660 6670 6680 6690 6700 6710 6720
ACACTTCCCTGGCCATCAGCCCAACAGACCAATCAAAAGCTGTACCACCTGCCCCACCCACCGTGGGCC 6720
CTCAGCAGCTTCCCTCACCTCCCGAGGATAGGACAGTCAAAAGACAGCACCCCAAGTTCTCTGGACCTAGA 6790
TCCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAAGCTGCAAACTACATGAGTCTCCAGATGAGAGAAAC 6860
ATCCTGGACCCCAAGCTTCAAGGAAACCTTGAAGGTTGGGCAATCACTGTCACCCCGGACACCCAGAC 6930
GCCTCCATCAGCTATCTGAGCTCTCTCTCCCTCTCTCTTTCAGAGCACCTGGCTCTCCAGCCCGAGC 7000
7010 7020 7030 7040 7050 7060 7070
AGGAGAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7070
CTAGGAAAGAAATGGTGGGGTGGGCTTGGGAACTTGTGCCCCCAAAACACATTAATGGCCCTCTCTAAT 7140
GAGCTTGGGGAAAGATGATCTTGGGCTTCTCCCTTGACTCTCTGTTCAATTAACAACTCTGGGCTTT 7210
CTGGGAGGAGGTTCAAGAAACATCAAAACCTGACAGCAGTTCGCCGAATTCAGCTTGGACTTAACCAAG 7280
CTGAAGTTGCTCAAAAGAGCCGAATTCAGGACAGCTGGGCTCCCATGGTATCTATCTGAGCTCCGCT 7350

Monday, 1 December 1997 14:12

lg pCB501

7360 7370 7380 7390 7400 7410 7420
ATGGGGCGCTGTCATCAGATCGCCATCTCGCGCCCGTGCCCTGACTTCFAAGGCCAATGACCTTCAAC 7420
ATCCCTACATGCTCTTTCGCCGIGGCCACCCCTATTTTGTATTATCMAAAACCTCTCTTA 7430
ATTTCTTTTGTGTTTTTACCTTCTTTTAACTCACCTCAACAATGAAATTGTGTAGATCAAAAAATAGAATT 7440
AATTCGTAATAAAAAAGTCGAAAAAATGTGCTCCCTCCCGCATTAATAAATTTCTATCCCAAAATCT 7450
ACACAATGTTCTGTGTACACITCTTAAGTTTTTTTACTTCTGATAAATTTTTTTGAAACATCATAGAA 7460
7710 7720 7730 7740 7750 7760 7770
AAAAACCGCACACAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTATTTCTTCGCA 7770
CGTCTGGGCTCTCATGACGTCAAATCATGCTCATCGTGAAAAAGTTTTGGAGTATTTTTGGAATTTTC 7780
AATCAAGTGAAAGTTTATGAAATTAATTTCTGCTTTTGTCTTTTGGGGTTTCCCTATTTGTTGTCA 7790
AGAGTTTCGAGGACGGGCTTTTCTTCTGCTAAATACAAAGTATTGATGAGCAGGATCAAGAAAGATCGG 7800
AAGAAGGTTTGGGTTTGGGCTCAGTGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCT 8050
8060 8070 8080 8090 8100 8110 8120
ATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAACATAACTGTTTCTACTAGTGG 8120
CGGTACGGGCGCTTTCTGCTCGCGGCTTTCGGTGATGACGCTGAAAACTCTGACACATGACAGCTCCCG 8130
AGACGGTCACAGCTTGCTGTAAGCGGATGCGGGGAGCAGACAAGCCGCTCAGGCGCGCTCAGCGGCTGT 8260
TGCGGGGTGTCGGGGCTGGCTTAACATATGCGGCATCAGAGCAGATTGTACGAGAGTGCACCAATATCGG 8330
TGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCAACAGGCGGCTTAAAGGCTCTGATACGC 8400
8410 8420 8430 8440 8450 8460 8470
CTATTTTATAGGTTAATGTCATGATAAATGTTTCTTACAGCTCAGGTGGCACTTTCTCGGGGAAATG 8470
TGGGGGAAACCTTATTTGTTTATTTCTTAAATACATTCAAAATATGATCGGCTCAAGAGACAAATACC 8540
CTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTAGAGTATCAACATTTCCTGTGCGGCTTATT 8610
TCTTTTTTGGGGCATTTTGCCTTCTGTTTTTCTCACCCAGAAACGCTGGTGAAAGTAAAGATGCTG 8680
AAGATCAGTTGGGTCCACGAGTGGGTTACATCGAAGTGGATCTCAACAGCGCTAAGATCTTGAAGATTT 8750
8760 8770 8780 8790 8800 8810 8820
TGGCCCCGAAGAAGCTTTTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGTGCGGCGGTAATACCG 8820
ATGACGCGGGCAAGAGCAACTCGGTCCCGCATACACATTTCTCAGAAATGACTTGGTTGAGTACTCAC 8890
CAGTCACAGAAACCATCTTACGGATTCATGACAGTAAGAGAATTATGCAAGTGGCTGCCATAACCATGAG 8960
TGATAACACTGCGGGCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAAACGCTTTTTCGAC 9030
AAGATGGGGGATCATGTAACGCGCTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG 9100
9110 9120 9130 9140 9150 9160 9170
AGGGTGACACCAGATGCGCTGAGGAAATGGCAACAAGCTTGGCGAAACTATTAACTGCGCAACGATGAC 9170
CTAGCTTCCCGGCAACAATTAATGAGACTGGATCGAGGGGGAATAAGTTCAGGACCACTTCTGCGCTCG 9240
GCGCTTCCCGCTGCTGCTTTATTGCTGATAAATCTGGAGCGGCTGAGGCTGCGCTCGCGGTAATGATG 9310
GAGCACTGGGGCAGATGGTAAAGCCCTCCGATGCTAGTATCTACACGACGGGGAGTCAGGCAACTAT 9380
GATGAACGAATAGACAGAATCGCTGATATAGGTGCTGCTGATTAAGCAATGGTAAGTTCAGAGCAAA 9450
9460 9470 9480 9490 9500 9510 9520
GTTTACTCATATATAGTTTGAATTAATTAACCTCAATTTTAAATTTAAAGGATCTGAGTGAAGATCT 9520
TTTTTAAATCTCATGAGCAAAAATGCTTAACCTGAGTTTCTGCTCAGCTCAGGCTCAGAGCCCGTAGA 9590
AAGAATCAAGGATCTTCTGAGATTCCTTTTCTGCGGCTAATCTGCTGCTTGAACAACAAAAACCA 9660
CGGCTACACCGGTTGTTGTTTCCCGCATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGCTTCA 9730
GAGAGCGGAGATACCAATACTGCTGCTTAGTGTAGCTGAGTGAAGGCAACGCTTCAAGAACCTGT 9800

Monday, 1 December 1997 14:12
fig pCB501

Page 5

9810 9820 9830 9840 9850 9860 9870
ASCACCGCCTACATACC TCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG3CGATAAGTCGTGT 9870
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCTGT 9870
GCACACAGCCAGCTTGGAGCGAAGGACCTACACCGAAGTGAAGTACCTACAGCGT3AGCATTGAGAAAG 10010
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGSTATCCGGTAAGCGGCAGGGTCG3AACAGGAGAGCGC 10010
ACGAGGGAGCTTCCAGGGGGAAACGCTGGTATCTTTATAGTCTGCGGGTTTCGCCACCCTGACTTG 10150
10160 10170 10180 10190 10200 10210 10220
AGCGTCGATTTTGTGATGCTCGTCA3CGGGGGCGGAGCCTATGGAAAAACGCCAGCAACCGCGGCCTTTT 10220
ACGGTTCCITGGCCTTTTGTGCTGCTCTTTGCTCACA TGTCTCTTCTGCGTTATCCCTGATCTGTGGAT 10290
AACCGTATTACCGCTTTTGTGAGTGTATACCGCTCGCCGCGAGCCGAACGACCGA3CGCAGCGAGTCAAG 10380
T3AGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCGCTCTCCCGCGCGTTGGCCGATTCAATTAATG 10430
CAGCTGGCACGACAGGTTTCCCGACT3GAAAGCGGGCAGTGAAGCGCAACGCAATTAATGTGAGTTAGCTC 10500
10510 10520 10530 10540 10550 10560 10570
ACTCATTAGGCACCCAGGCTTACAGTTTATGCTTCCGGCTCGTATGTTGTGTGGAATGTGAGCGGA 10570
AACAAATTCACACAGGAAACAGCT 10594

SEQUENCE ID No 10

Page

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence

Enzymes: 100 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

fig 13 page

GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAACTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT 100
CTGCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG 200
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAAC TGCGTGTAAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGG
G G R . V V R E Q N L S Y N K A R L D R O L H E E S A . G . A F C

CTGCTTCGGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA 300
GACGAAGCGCTACATGCCCGTCTATATGCGCAACTGTAAC TAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
A A S R C T G Q I Y A L T L I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCGGGTTACATAACTTACGGTAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT 400
ACCTCAAGGCGCAATGATTGAATGCCATTACGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAAC TGCAAGTATTACGTCATACAAGGGTATCA
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGACTATTACGGTAACCTGCCCACTTGCGCAGTACATCAAGTGTATCATATGCCAAGTACGCC 500
TTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAAATGCCATTGACGGGTGAACCGTCAATGATGTTACATAGTATACGGTTATCGCGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGGTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA 600
GGATAACTGCAAGTACTGCCATTACCGGGCGGACCGTAATACGGGTGATGACTGGAATACCTGAAAGGATGAACCGTCAATGATGATGATCAATCAG
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGATSCGTTTTGGCAGTACATCAATGGCGGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAA 700
AGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAAC TGAGTGCCCTAAAGGTTCAAGGTTGGGGTAAC TGCAAGT
R Y Y H G D A V L A V P Q W A V I A V . L T G I S K S P P H . R Q

TGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTAACAAC TCCGCCCCATTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAG 800
ACCTCAAAACAAAACCGTGGTTTTASTTGCCCTGAAAGGTTTTACAGCATGTTGAGGCGGGTAAC TGCTTTACCGGTCATCCGACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTAC TGCGTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGTGGCTAGC 900
CAGATATATTGCTCTGAGAGACCGATTGATCTCTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGACCGGATCG
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A

GTTTAAACTTAAGCTTACCATGGGSGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACGCAAAATGGGTGGGATCTGTACGAC 1000
CAAAATTTGAATTCGAATGGTACCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATGCTACTGACCACTGTCTGTTTACCCAGCCCTAGACATGCTG
F K L K L T M G S S H - H H H H G M A S M T S G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GATGACGATAAGGTACCTAGGATCCATATGCCCTCCCTGCGCGTGCAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCSTCGACA
CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACASTTATGTATAGTCAGAGGGAGTTCCAGACTTCTCTTTACGCAGCTG
pCB201 insert = U4
U4 ORF
D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C V D
GGCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCTCGTCTCTCGGGCCCCAGCGGCAC
CGGACCACAAGCTCTGCGACTAGGGGTTGCGCTACTACGTGCTGATGATTTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCCGGGGTCGCCGT
pCB201 insert = U4
U4 ORF
S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S G P S G T
GGGCAAGACCTACCTGACCAATCGCTTGCGCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCA
CCCGTTC TGGATGGAC TGGTTAGCGAACC GGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGT
pCB201 insert = U4
U4 ORF
G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q
CACTCTTGAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAAGATAGACCGGGAACAGGAATTGGGGATGTGCCCC TGGTGATTCTATTGATG
GTGAGAACGTTCTTAGAGCTTGACATAGAAAGGTTGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCCTAAGATAACCTAC
pCB201 insert = U4
U4 ORF
S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V I L L D
AGCTGAGTGAAGCAGGCTCCATCAGTGAAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCC TATATTATAGGTACCAACATCAGCTGT
TGGACTCACTTCGTCCGAGGTAGTCACTCAACCAAGTACCCCGGAGTGGACGTTTATAGTATTACAGGATATAATATCCATGGTGGT TAGTGGAC
pCB201 insert = U4
U4 ORF
D S E A G S I S E L V A G A L T C K Y H K C P Y I I G T T N Q P A
AAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCTGGTTGGTTACCTGAGG
TTTTTACTGTGGGTGGGTACCGAAGTGAAGTCAAGTCTTACAACGTGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGAGTCC
pCB201 insert = U4
U4 ORF
K M T P N H G L H L S F R M L T F S N N V E P A N G F L V R Y L F

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTSGTATCATCTCCACACE"
TCCTTCGACCATCTCACTCTGTCGCTGTAGTTACGGTTGTTCTTCGACGAAGCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGA 700

pCB201 insert = U4

U4 ORF

R K L V E S D S D I N A N K E E L L R V L D V V P K L V Y H L H T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGTCTTCTTGTGCTGTCCATTGGCATTGAGGACTTCGGGACCTGGTTCATTGACCT
AGGAACCTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCCTGGACCAAGTAATGGA 800

pCB201 insert = U4

U4 ORF

F L E K H S T S D F L I G P C F F L S C P I G I E D F R T V F I D L

GTGGAACAACCTCTATCATTCCCTATCTACAGGAAGGAGCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTTGGGAGGACCCAGTGGAAATGGGT
CACCTTGTTGAGATAGTAAGGGATAGATGTCTTCTCGGTTCTTACCTATTTCAGGTACCTGTCTTCGACGAACCCCTCTGGGTACCTTACCCAG 900

pCB201 insert = U4

U4 ORF

V N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D P V E V V

CGGGACACACTTCCTTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCACCGTGGGCCCTCACAGCATTGCTCAGCTCCG
GCCCTGTGTGAAGGGACCGGTASTCGGGTTGTTCTGTTAGTTTCGACATGTTGGACGGGGGTGGGTGGCACCCGGGAGTGTCTGAACGGAGTGAAGGG 1000

pCB201 insert = U4

U4 ORF

R D T L P V P S A Q Q D Q S K L Y H L P P P T V G P H S I A S P P

AGGATAGGACAGTCAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAATACATTGAGTC
TCCTATCTGTCAGTTTCTGTGCTGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACCTG 1100

pCB201 insert = U4

U4 ORF

E D R T V K D S T P S S L D S D P L M A M L L K L Q E A A N Y I E S

TCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTACCCCCGGACAGCAGAACGCTGCATCAGTA
AGGTCTAGCTCTTTGGTAGGACCTTGGGTTGGAAGTCCGTTGTGAAATTCCTAAGCCGTTAGTGACAGTGGGGGCTGTGCTCTTGCACCGTACTGGA 1200

pCB201 insert = U4

U4 ORF

P D R E T I L C P N L Q A T L G F G N H C H P R T A E R V H Q L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

TCTTAGCTCCTCCTCTCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGT
AGAAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGCGGGTCTCTCTTGTCTCCTCCTCTACTTTCTCTCCTGTCCAA 2300

pCB201 insert = U4

S L L L S P L L F Q S T G S P A P G G E Q E G G G D E R G G T G

CTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTGTGCCCTTAAACACATTTACTGGCCTCCTCTAATGACT
GAACACGACATGGAACCTCTGAAGGATCTCTTACCACCCACCGCAAAACCTTGAACACGGGGGATTGTGTAAATGACCGGAGGAGATTACTGA 2400

pCB201 insert = U4

S W C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TGCGGAAAAGATGATTCTGGGTCTTCCCTTGACTTCTGTITTAACAACTCTGGGCTTTCTGGGGAGGGGTTTCAGAAAACATCAAAACACTGC
AACCCCTTTTCTACTAAGACCCAGAAAGGGAAGTGAAGAACAAGTTAATGTTTGGAGACCCGAAAGACCCCTCCCAAGTCTTTTGTAGTTTGTGACG 2500

pCB201 insert = U4

V G K D D S G S F P . L L V S I T N S V A F V G G V Q K T S K H C

AGCAGTTCTAAATGATTCTCACAAGCAACCTGAGAGAGACAGTCTTGTGAGGAGATCTGGGGAGGCAGGAAGCTCTCAGATTTTCTCACAGACCC
TCGTCAAGGATTTACTAAGAGTGTTCGTGGGACTCTCTCTGTGAGAACACTCCCTCTAGACCCCTCCGTCTTCGAGGAGTCTAAAAGAGTGTCTGGG 2600

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I W G R Q E A P O I F S Q T

TTCCCAATTCATCACCCTGCAACACTCGTCCGGAATCTGTCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG
AAGGGTTAAGGTAGTGTGACGGTTGTGAGCAGCCCTTAAGAGCTCTATAGTCTGTACCCGCCGGCGAGCTCAGATCTCCCGGGCAATTTGGGCGAC 2700

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q V R P L E S R G P V . T R .

ATCAGCTCGACTGTGCCTTCTAGTTGCCAGCATCTGTGTGTTGCCCTCCCCGTGCTTCTTGACCTGGAAGGTGCCACTCCCACTGCTCTTCC
TAGTCGGAGCTGACACGGAAGATCAACGGTCCGTAGACAACAACGGGGAGGGGACCGAAGGAACCTGGGACCTTCCAGGTGAGGGTGACAGGAAGG 2800

S A S T V P S S C Q P S V V C P S P V P S L T L E G A T P T V L S

TAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCTATTCTCTGTTGGGGTGGGGTGGGGCAGGACAGCAAGGGGAGGATTGGGAAGACA
ATTATTTTACTCCTTTAAGTAGCGTAACAGACTCATECACAGTAAGATAAGACCCCAACCCCAACCCGTCCTGTCTGTTCCCTCTCTAACCCTTCTGT 2900

N E E I A S H C L S R C H S I L G G G V G Q D S K G E D V E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGTATCCCCACGCGCCCTGTAGCGGGC
TATCGTCCGTACGACCCCTACGCCACCGGAGATACGGAAGACTCCGCTTCTTGTGTCGACCCCGAGATCCCCATAGGGGTGCGCGGGACATCGCCGCG 3000

N S R H A G D A V G S M A S E A E R T S V G S R G Y P H A P C S G A

ATTAAGCGCGGGGTGTGGTGGTTACGCGCAGCTGACCGCTACACTTGCAGCGCCCTAGCGCCGCTCCTTTCTGCTTCTTCTCTCTCTCTCTGCG
TAATTCGCGCGCCACACCAATGCGGCTCCTCTGCGGATGTGAACGCTCGCGGGATCGCGGGCAGGAAAGCGAAAGAGGAAAGAGGAGGAGG 3100

L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACGTTCCCGGGCTTTCCCGCTCAAGCTCTAAATCGGGCATCCCTTTAGGTTCCGATTAGTGTCTTACGGCACCTCGACCCCAAAAACTTGATTAGG
TGCAAGCGGGCGAAAGGGGAGTTCGAGATTAGCCCGTAGGGAAATCCCAGGCTAAATCAGGAATGCCGTGGAGCTGGGGTTTTTGAACTAATCC 3200

T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D .

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1>5082) Site and Sequence

Page

GTGATGGTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTGTTCCAACTGG
CACTACCAAGTGCATCACCCGGTAGCGGACTATCTGCCAAAAAGCGGAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGTTTGACC 330
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G
AACAACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGGGATTTCGGCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTT
TTGTTGTGAGTTGGGATAGAGCCAGATAAGAAACTAAATATTCCTTAAACCCCTAAAGCGGATAACCAATTTTTACTCGACTAAATGTTTTTAA 340
T T L N P I S V Y S F D L . G I L G I S A Y V L K N E L I . Q K F
AACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGAAAGTCCCGAGGCTCCCGAGGAGGAGAGTATGCAAGCATGCATCTCAATTAGT
TTGCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTGAGGGTCCGAGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCA 350
N A N . F C G M C V S . G V E S P Q A P Q A G R S H Q S M H L N .
CAGCAACCGAGTGTGGAAGTCCCGAGGCTCCCGAGGAGGAGTATGCAAGCATGCATCTCAATTAGTCAAGCAACCATAGTCCCGCCCTAACTCC
GTCGTTGTTCCACACCTTTGAGGGTCCGAGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAG 360
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P
GCCCCCGCCCTAACTCCGCCAGTTCGCCCTATTCGCCCTATGCTGACTAATTTTTTTATTTATGCAAGGCGGAGGCGCCCTCTGCCTCT
CGGGTAGGGCGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCAGTATTAATAAATAAATACGTCCTCCGGCTCCGGCGGAGAGGAG 370
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S
GAGCTATTCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAGCTCCCGGAGCTTGATATCCATTTTCGGATCTGATCAAGAGA
CTCGATAAGGTCTTCATCACTCTCCGAAAAACCTCCGGATCCGAAACGTTTTTCGAGGGCCCTCGAACATATAGGTAAAGCCTAGACTAGTTCTCT 380
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E
CAGGATGAGGATCGTTTCGATGATTGAACAAGTGGATTGACGCGAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATCGGCTATGACTGGGCACAACA
GTCTACTCTAGCAAGCGTACTAATCTGTTCTACCTAACGTGCGTCCAAGAGGCGGGCGAACCCACCTCTCCGATAAGCCGATAGTACCCGTTG 390
T G . G S F R M I E Q D G L H A G S P A A W V E R L F G Y D W A Q Q
GACAACTGGCTGCTGATGCCCGGTGTTCCGGCTGTCAGCGAGGGGCGCCCGTTCTTTTGTCAAGACCGACCTGTCGGGTGCCCTGAATGAAGT
CTGTAGCGAGGAGTACGGCGGCACAAGGCGSACAGTCCGCTCCCGGGGCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGAG 400
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L
CAGGACGAGGAGCGGGCTATCGTGGCTGGCCACGAGCGGGCTTCCTTGGCGAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTA
GTCTGCTCCGTCGCGCGATAGCACCGACCGGCTGCTGCCCGCAAGGAACGCTCGACACGAGCTGCAACAGTGAATTCGCCCTTCCTTGACCGAGTA 410
G D E A A R L S V L A T T G V P C A A V L D V V T E A G R D V L L
TGGCGAAGTGGCGGGCAGGATCTCTGTATCTCACCTTGCTCTGCGGAGAAAGTATCCATCATGGCTGATGCAATGGCGGGCTGCATAGGCTTGA
ACCGGCTTACGGGCCCCGTCTAGAGGACAGTASAGTGAACGAGGAGCGCTCTTTCATAGGTAGTACCGACTACGTTACGCCCGGACGATATGCGAAG 420
L S E V P G Q D L L S S H L A P A E K V S I M A D A M R R L H T L D
TCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAA
AGCGCGATGGAGCGGTAAAGCTGGTGGTTGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCTACCTTCGGCCAGAACAGCTAGTCTTACTAGACCTGCT 430
P A T C P F D H Q A K H R I E R A R T R M E A G L V D Q D D L D E
GAGCATCAGGGGCTCGCGCCAGCGCAACTGTTCTCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTGTGACCATGSCGATSCCTGCTTGG
CTCGTAGTCCCGAGCGCGGTGGCTTGACAAGCGGTCCGAGTTCGCGCGTACGGGCTGCGCTCTAGAGCAGCACTGGGTACCGCTACGGAGCAAG 440
E H G G L A P A E L F A R L K A R M P D G E D L V V T H G D A C L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

CGAATATCATGGTGGAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTACCGCGGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCGTATCGCAACCGATGGGCACT 500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D
TATTGCTGAAGAGCTTGGCGGCAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACC GCCGCTTACCGGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 500
I A E E L G G E V A D R F L V L Y G I A A P D S Q R I A F Y R L L
GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTTCGATTCCACCGCCGCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 500
D E F F . A G L V G S K . P T K R R P T C H H E I S I P P P P S M
AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCCACCCCAACTTGTTTATT
TTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 500
K G V A S E S F S G T P A G . S S S A G I S C V S S S P T P T C L L
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACCTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
CGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTAAAGTGTATTTTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGT 500
O L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S
ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTCTGTGTGAAATTGTTATCCGCTCACAAT
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACC GCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTA 500
M Y L I M S V Y R R P L A R A W R N H G H S C F L C E I V I R S Q
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG 5082
AGGTGTGTGTATGCTCGGCCCTTCGTATTTACATTTTCGGACCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAAC
F H T T Y E P E A . S V K P G V P N E . A N S H . L R V

Claims

1. A vertebrate protein homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein
5 comprises an amino acid sequence having a statistically significant homology to the amino acid sequence of said UNC-53 protein of C. elegans illustrated in Figure 2.
- 10 2. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D or E as illustrated in Figure 9a, or
15 block F in Figure 12a or a sequence having a statistically significant homology therewith.
- 20 3. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E or F which differ from those blocks of Figure 9a or 12a only in conservative amino acid changes.
- 25 4. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown from nucleotide positions 1 to 6013 illustrated in Sequence ID No. 3.
- 30 5. A vertebrate protein comprising an amino acid sequence which comprises one or more of the prosite signatures as illustrated in Figure 28 for each of said sequences of homology as claimed in claim 2.
- 35 6. A vertebrate protein comprising an amino acid

- 185 -

sequence as claimed in any one of claims 1 to 6 which is a human protein or a mouse protein.

5 7. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown in Sequence ID No. 4.

10 8. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 in one or more conservative amino acid changes.

15 9. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 2 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 in one or more
20 conservative amino acid changes.

25 10. A cDNA encoding a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 9.

30 11. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 only in one or more conservative amino acid changes.

35 12. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 2 or an amino

- 186 -

acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 only in one or more conservative amino acid changes.

- 5 13. A cDNA according to any of claims 10 or 11 which cDNA comprises at least from nucleotide position 1 to position 6013 of the sequence as shown in Sequence ID No. 3.
- 10 14. A cDNA according to claim 10 or 12 which comprises the nucleotide sequence illustrated in Sequence ID No. 4.
- 15 15. A nucleic acid molecule capable of hybridising to the DNA sequences according to any of claims 10 to 14 under high stringency conditions.
- 20 16. A DNA expression vector which comprises a cDNA as claimed any of claims 10 to 14.
- 25 17. A vector according to claim 16 which comprises a promoter of C. elegans UNC-53 protein or a vertebrate homologue thereof according to any of claims 1 to 9.
- 30 18. A vector according to claim 17 wherein said promoter sequence is derived from a gene encoding a mouse or human homologue of an UNC-53 protein of C. elegans.
- 35 19. A vector according to any of claims 16 to 18 which further comprises a sequence encoding a reporter molecule.
20. A vector according to claim 19 wherein said

- 187 -

reporter molecule is a fluorophore.

21. A host cell transformed or transfected with the vector of any of claims 16 to 20.

5

22. A host cell transformed or transfected with the vector of claims 19 or 20.

23. A host cell according to claims 21 or 22, which cell comprises a prokaryotic cell such as a bacterial cell or a eukaryotic cell such as a fungal, an animal, a plant or an insect cell.

10

24. A transgenic cell, tissue or organism comprising a transgene capable of expressing a protein according to any of claims 1 to 9.

15

25. A transgenic cell, tissue or organism according to claim 24 which comprises any of a COS cell, Hep G2, MCF-7 cell, N4 mouse neuroblastoma cell, a NIH3T3 cell, or colorectal carcinoma or human derived cells.

20

26. A transgenic cell, tissue or organism according to claim 24 or 25 wherein said transgene comprises a vector according to any of claims 16 to 20.

25

27. A transgenic cell, tissue or organism according to claim 24 to 26 wherein said transgene comprises a vector according to claim 19 or 20.

30

28. A transgenic cell, tissue or organism according to any of claims 24 to 26 wherein said organism comprises any of an insect, a fungus, a non-

35

human mammal, a plant or a nematode worm.

29. A method of producing a mutant vertebrate non-human organism which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration, which method comprises inducing a mutation in the wild type gene encoding the vertebrate homologue of an UNC-53 C. elegans protein.

30. A vertebrate protein homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, for use as a medicament to promote neuronal regeneration, revascularisation, wound healing or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

31. A vertebrate protein homologue of an UNC-53 protein of C. elegans for use as claimed in claim 30 wherein said vertebrate human homologue is as claimed in any one of claims 1 to 9.

32. Use of a vertebrate protein homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation, wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

33. Use of a vertebrate protein homologue of UNC-53 protein of C. elegans according to claim 32 wherein said vertebrate protein homologue is as

claimed in any one of claims 1 to 9.

5 34. A pharmaceutical composition comprising a vertebrate homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate protein, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

10 35. A pharmaceutical composition as claimed in claim 34 which comprises a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 9.

15 36. A nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, for use as a medicament.

20 37. A nucleic acid sequence according to claim 36 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said cDNA sequence.

25 38. Use of a nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, in the manufacture of a medicament to
30 promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

35

39. Use of a nucleic acid sequence according to claim 38 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said nucleic acid sequence.

5

40. A pharmaceutical composition comprising a nucleic acid sequence according to claim 36 or 37 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

10

41. A pharmaceutical composition according to claim 40 wherein said nucleic acid sequence is a cDNA sequence as claimed in any of claims 10 to 14.

15

42. A method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, cell shape or motility or the direction of cell migration, which method comprises contacting said compound with a host cell according to claim 21 or 23 or a transgenic cell as claimed in any of claims 24 to 27 and screening for a phenotypic change in said cell.

20

43. A method according to claim 41 which is capable of determining whether said compound is an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which said vertebrate homologue of an UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue is a component or is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell.

25

30

35

44. A method according to claim 43 wherein said

- 191 -

method is capable of determining whether said compound is an inhibitor or an enhancer of said vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

45. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in cell growth, or shape or a change in cell motility.

46. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in filopodia outgrowth, ruffling behaviour, cell adhesion, contact inhibition or the length of neurite growth.

47. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an N4 neuroblastoma cell and the phenotypic change to be screened is the length of neurite growth.

48. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an MCF-7 breast carcinoma cell or an NIH3T3 cell and the phenotypic change to be screened is the extent of phagokinesis or contact inhibition.

49. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of cell shape, cell growth or motility or of the direction of cell migration, which method comprises administering said compound to a transgenic organism according to any of claims 24 to 28 or a mutant organism produced according to the method of claim 29

and screening for a phenotypic change in said organism.

50. A method according to claim 49, wherein said
5 method is capable of determining whether said compound
is an inhibitor or enhancer of a protein of the signal
transduction pathway of said transgenic or mutant
organisms, of which the vertebrate homologue of UNC-53
protein of C. elegans or a functional equivalent,
10 derivative, fragment or bioprecursor of said
vertebrate homologue is a component, or is an
inhibitor or an enhancer of a parallel or redundant
signal transduction pathway in said cell.

51. A method according to claim 50 wherein said
15 method is capable of determining whether said compound
is an inhibitor or an enhancer of the vertebrate
homologue of UNC-53 protein itself or a functional
equivalent, fragment, derivative or bioprecursor of
20 said vertebrate homologue.

52. A compound which is identifiable by the
method according to any one of the claims 42 to 51 as
an enhancer of the regulation of cell shape, or growth
25 or motility or the direction of cell migration for use
as a medicament for promoting neuronal regeneration,
revascularisation or wound healing or for treatment of
chronic neurodegenerative diseases or acute traumatic
injuries or fibrotic disease.

53. Use of a compound which is identifiable by
30 the method according to any one of the claims 42 to 51
as an enhancer of the regulation of cell shape, or
growth or motility or the direction of cell migration
35 in the preparation of medicament for promoting

neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

5

54. A pharmaceutical composition comprising a compound identified according to the method of any of claims 42 to 51 claim and a pharmaceutically acceptable carrier, diluent or excipient therefor.

10

55. A compound which is identifiable by the method according to any one of claims 42 to 51 as an inhibitor of the regulation of cell motility, growth, or shape, or the direction of cell migration, for use as a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

15

56. Use of a compound according to claim 55 in the manufacture of a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

20

57. A pharmaceutical composition comprising the compound as claimed in claim 55, and a pharmaceutically acceptable carrier diluent or excipient therefor.

25

58. A method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene encoding a vertebrate homologue of UNC-53 protein of C. elegans, which method comprises the steps of (a) contacting said compound with a cell according to any of claims 22 or 27 and (b) monitoring the level of said reporter molecule and comparing the results

30

35

obtained from said monitoring step with a control comprising a cell according to claims 22 or 27, which cell has not been contacted with said compound.

5 59. A method as claimed in claim 58 wherein said reporter molecule detected is mRNA or green fluorescent protein.

10 60. A compound which is identifiable by the method according to claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate
homologue of an UNC-53 protein of C. elegans or a
functional fragment of said gene, for use in promoting
15 neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-
degenerative diseases or acute traumatic injuries or
fibrotic disease.

20 61. Use of a compound which is identifiable by the method of claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate
homologue of an UNC-53 protein of C. elegans or a
functional fragment of said gene, in the manufacture
25 of a medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment
of chronic neuro-degenerative diseases or acute
traumatic injuries or fibrotic disease.

30 62. A pharmaceutical composition which comprises the compound of claim 60 and a pharmaceutically
acceptable carrier, diluent or excipient therefor.

35 63. A compound which is identifiable by the method of claims 58 or 59 as an inhibitor of
transcription of a gene coding for a vertebrate

homologue of a UNC-53 protein of C. elegans or a functional fragment of said gene for use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

5

64. Use of a compound which is identifiable by the method of claims 58 or 59 as an inhibitor of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for alleviating spread of disease inducing cells or metastasis or loss of contact inhibition.

10

65. A pharmaceutical composition which comprises the compound of claim 63 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

15

66. A kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell motility, growth or shape or the direction of cell migration which kit comprises at least one transgenic cell as claimed in any one of claims 22 to 25 to be contacted with said compound and at least one cell according to claims 21 to 28 to be used as a control and means for contacting said compound with one of said at least one transgenic cells.

20

25

67. A kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene which kit comprises at least one cell as claimed in any one of claims 21 to 25 means for contacting said compound with said cells.

30

35

- 196 -

68. A kit for determining whether a compound is an enhancer or an inhibitor of the activity of a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue protein, which kit comprises at least, one vertebrate mutant non-human organism produced according to the method as claimed in claim 29 or a transgenic organism as claimed in claims 24 to 28 and a wild type of said vertebrate mutant organism.

69. A method of identifying vertebrate homologues of an unc-53 gene of C. elegans or a functional fragment thereof, which method comprises hybridizing to a DNA library a suitable oligonucleotide sequence of between 15 to 50 nucleotides of the nucleic acid sequence encoding unc-53 or a functional equivalent, derivative or bioprecursor thereof, under appropriate conditions of stringency to identify genes having statistically significant homology with the cDNA according to any of claims 10 to 14.

70. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said vertebrate homologue is a component, which method comprises:

(a) contacting an extract of said cell with an antibody to the vertebrate homologue of the UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said protein,

(b) identifying the antibody/vertebrate

- 197 -

homologue complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein of C. elegans other than the antibody.

71. A method of identifying a further protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein or a functional equivalent, fragment or bioprecursor of said UNC-53 protein is a component, which method comprises:

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein of C. elegans in claim 70,

(b) contacting a cell extract with said antibody and identifying the antibody/protein complex,

(c) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(d) optionally repeating steps (a) to (c) to identify further proteins in said pathway.

72. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue protein is a component, which method comprises

(a) contacting an extract of said cell with the vertebrate homologue of an UNC-53 protein of C. elegans or a functional

- 198 -

equivalent, derivative or bioprecursor of said vertebrate homologue,

(b) identifying any vertebrate homologue of UNC-53 protein/protein complex formed and

5 (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

10 73. A method according to claim 72 which further comprises contacting a cell extract with any protein identified from step (c) not being the same as the vertebrate homologue of UNC-53 protein used and
15 repeating steps (b) and (c) so as to identify any further protein involved in the signal transduction pathway of said cell.

20 74. A method of identifying a protein involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans is a component which method comprises:

25 (a) providing an appropriate host cell having a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain,
(b) expressing in said host cell a first hybrid DNA sequence encoding a first fusion
30 of a fragment or all of a DNA sequence according to any of claims 10 to 14 and either said DNA binding domain or the activating domain of the transcription factor,

35 (c) expressing in the host cell at least one second hybrid DNA sequence encoding a

- 199 -

putative binding protein to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion,

5 (d) detecting any binding of the protein being investigated with a protein according to any of claims 1 to 9 by detecting for the production of any reporter gene product in said host cell.

10

75. A protein identified by the method of any one of claims 70 to 74 for use as a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-
15 degenerative diseases or acute traumatic injuries or fibrotic disease.

76. Use of a protein identified by the methods of any one of claims 70 to 74 in the manufacture of a
20 medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

25 77. A pharmaceutical composition comprising a protein identified by the methods of any one of claims 70 to 74 and a pharmaceutically acceptable carrier, diluent, or excipient therefor.

30 78. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent fragment, derivative or bioprecursor of said vertebrate homologue which process comprises culturing the cells of any of claims
35 21 to 28 and recovering said vertebrate homologue of

- 200 -

UNC-53 protein expressed.

79. A process for producing a vertebrate
homologue of an UNC-53 protein of C. elegans or a
5 functional equivalent, fragment, derivative or
bioprecursor of said protein which process comprises
culturing an insect cell transfected with a
recombinant Baculovirus vector, said vector comprising
a DNA insert encoding said vertebrate homologue of
10 UNC-53 protein or a functional equivalent, fragment or
bioprecursor of said vertebrate homologue, downstream
of the Baculovirus polyhedrin promoter, and recovering
the expressed vertebrate homologue of UNC-53 protein.

15 80. A nucleotide sequence comprising the
sequence as shown in figure 15.

81. A nucleotide sequence comprising the
sequence as shown in figure 16.

20 82. A nucleotide sequence comprising the
sequence as shown in figure 17.

83. A method of detecting whether a compound is
25 an inhibitor or an enhancer of expression of a
vertebrate homologue of an UNC-53 of C. elegans, or a
functional equivalent, derivative or fragment of said
vertebrate homologue which method comprises contacting
a cell expressing said homologue with said compound
30 and monitoring for a phenotypic change compared to a
control cell which has not been contacted with said
compound.

84. A method according to claim 83 wherein said
35 cell comprises a cell according to any of claims 21 to

- 201 -

28.

85. A method according to claim 83 wherein said cell has undergone loss of contact inhibition.

5

86. A method according to any of claims 83 to 85 which is capable of determining whether said compound is an inhibitor of expression of said vertebrate homologue in which the compound to be tested comprises a nucleic acid.

10

87. A method according to claim 86 wherein said nucleic acid sequence comprises an antisense DNA or RNA sequence.

15

88. A method according to claim 87 wherein said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

20

89. A method according to any of claims 83 to 85 wherein said compound to be tested comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue.

25

90. A method according to claim 89 wherein said protein comprises a protein identified according to any of the methods of claims 70 to 74.

30

91. A pharmaceutical composition comprising a compound identified according to any of claims 83 to 89 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

35

92. A nucleic acid sequence identified according to the method of any of claims 86 to 88 for use in

- 202 -

treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor thereof.

93. Use of a nucleotide sequence identified according to the method of any one of claims 86 to 88 in the preparation of a medicament for the treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

94. A nucleic acid according to claim 92 for use in the preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans.

95. A NIH3T3 cell line transfected with pcB201 and deposited under LMBP Accession No. 1603CB.

96. A plasmid pcB 201 of Sequence ID No. 10 deposited under LMBP Accession No. LMBP 3594.

97. A MCF-7 cell line transfected with plasmid pcB 201 deposited under LMBP Accession No. LMBP 1601CB.

98. An assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, or a functional equivalent,

- 203 -

derivative or bioprecursor thereof, which antibody is linked to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

5

99. An assay according to claim 98 wherein said reporter molecule is an antibody conjugated with a suitable fluorophore or detectable enzyme.

10

100. A method for detecting for expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue or a functional equivalent, fragment or bioprecursor therefor with a cell extract which probe is linked to a reporter and analysing for the presence of said reporter.

20

101. A method according to claim 100 wherein said probe comprises a complimentary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

25

102. A method according to claim 101 wherein said complimentary sequence is a 3' or 5' untranslated region of said mRNA.

30

103. A method according to claims 100 or 102 wherein said reporter comprises a radiolabel.

35

- 204 -

104. A method according to claim 100 wherein
said probe comprises an antibody specific for said
vertebrate homologue of said UNC-53 protein or a
functional equivalent, derivative, fragment or
5 bioprecursor therefor.

105. A method according to claim 104 wherein
said reporter comprises an antibody conjugated with a
detectable fluorophore or enzyme.
10

106. Phage Lambda clone 3b of Sequence ID No. 5
deposited under Accession No. LMBP 3595.

107. A method of determining whether a compound
15 is an inhibitor or an enhancer of association of UNC-
53 or a vertebrate homologue thereof according to any
of claims to 1 to 9 to microtubules or plus end
regions thereof, which method comprises:-

(a) contacting said compound with a
20 transgenic cell, tissue or organism
expressing UNC-53 protein or said vertebrate
homologue and which protein is operably
linked to a reporter molecule.

(b) screening for the localisation of said
25 reporter molecule as compared to a cell
according to step (a) which has not been
contacted with said compound.

108. A compound identifiable by the method
30 according to claim 107.

109. A compound identifiable by the method
according to claim 107 as an inhibitor of localisation
or association of UNC-53 or said vertebrate homologue
35 with microtubules or the plus end region thereof for

- 205 -

use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

110. A compound identifiable by the method
5 according to claim 107 as an enhancer of association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof, for use in promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic
10 neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

111. A pharmaceutical composition comprising the compound according to claims 108 or 109 and a
15 pharmaceutically acceptable carrier, diluent or excipient therefor.

112. A kit for determining whether a compound is an inhibitor or an enhancer of association of UNC-53
20 or a vertebrate homologue thereof according to any of claims 1 to 9 with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing UNC-53 and a reporter molecule or a cell according to any of claims 20 to 24
25 and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells.

113. A composition comprising UNC-53 of C. elegans
30 elegans or a vertebrate homologue thereof according to any of claims 1 to 9 linked to a compound identified as an inhibitor or enhancer of association of UNC-53 or said vertebrate homologue with microtubules or their plus end regions for use in targeting said
35 compound to said microtubule or the plus end regions

- 206 -

thereof.

114. A composition according to claim 113 which further comprises a cell transformation or
5 transfecting agent.

115. A method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue
10 or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed
15 and which results in targeting said protein to said microtubule or a plus end region thereof.

116. A method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue
20 thereof according to any of claims 1 to 9, which method comprises

a) contacting either an extract from a cell expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising candidate UNC-53
25 modifying enzymes in the presence of an indicator of covalent modification of a protein,

b) identifying any covalently modified UNC-53 protein from step a),

c) identifying said molecule involved in said
30 modification step.

117. A method according to claim 112, wherein said indicator comprises ³²P.

35

- 207 -

118. A method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which method comprises contacting said
5 compound with a cell, tissue or organism according to claim 27, and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.

10 119. Plasmid pLM1 of Sequence ID No. 6 deposited under Accession No. LMBP 3762.

120. Plasmid pLM4 of Sequence ID No. 7 deposited under Accession No. LMBP 3763.

15

121. Plasmid pEGF72 of Sequence ID No. 8 deposited under Accession No. LMBP 3764.

122. Plasmid pCB501 of Sequence ID No. 9
20 deposited under LMBP Accession No. LMBP 3765.

123. A worm strain comprising a chimeric C.elegans human unc-53 gene deposited under LMBP Accession No. LMBP-1663CB.

25

124. A vertebrate homologue according to any of claims 1 to 3 which is a mouse homologue.

125. A homologue according to claim 125 having
30 the sequence illustrated in Figure 14.

35

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 16

AlwI
ACCTCGCTCTGCTAATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA 9300
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCATTCCGCGT
P R S A N P V T S G C C Q W R . V V S Y R V G L K T I V T G . G A

Apal I
GCGGTCGGGCTGAACGGGGGTTCTGTCACACAGCCAGCTTGGAGCGAACGACCTACACGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC 9400
CGCCAGCCCGACTTGCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCGCGG
A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K R

ACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATC 9500
TGCGAAGGGCTTCCCTCTTTCGCTGTCCATAGGCCATTGCGCGTCCGAGCCTTGCTCTCGCGTGCTCCCTCGAAGGTCCCTCTTTCGCGGACCATAG
H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V S

TTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC 9600
AAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACCTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTTGCGGTCGTTGCGCGG
L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R G

Ava III
Nsi I
CTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 9697
GAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence

Enzymes: 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTACGGTAAATGGCCCGCTGGGTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCACTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTTACGGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACACGCGGCGGACCGTATAGTTACCGGACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H O V A V
TAGCGGTTTGACTCAGGGGATTTCAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCAAACTGAGTGCCCTAAAGGTTAGAGGTGGGGTAACGTCAGTTACCTCAAACAAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCA
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S
ACAACTCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTGTCAGGTGGGAGGTCATATAAGCAGAGCTGGTTTGTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAACGCGTTTACCGCCATCCGCACATGCCACCTCCAGATATATTCGTCGACCAAACTACTGGCAGTCTAGGCGATCGCGAT
O L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAACGCGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACCTCGTTCCCGCTCCTCGACAAGTGGGCCACACGGGTAGGACCAGCTCGACCTGCCGCTGCATTGCGGGTGTCAAGTGGC
eGFPc.e.unc53sma
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTTATCTGCACCACCGGCAAGCTGCCGCTGCCCTGGCCACCCCTCGTGAC
ACAGGCGCTCCCGCTCCCGCTACGGTGATGCCGTTGACTGGGACTTCAAGTAGACGTGGTGGCCGTTTCGACGGGACGGGACCGGGTGGGAGCACTG
eGFPc.e.unc53sma
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T
CACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCATGAAGCAGCAGACTTCTTCAAGTCCGCGATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTCTGAAGAAGTTACGGCGGTACGGGCTTCCGATGCAGGTCCTC
eGFPc.e.unc53sma
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGSCATCG
GCGTGGTAGAAGAAGTTCTTGCTGCCGTTGATGTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTASCTGACTTCCCGTAGC
eGFPc.e.unc53sma
P T I F F K D D G N Y K T R A E V K F E G D I L V N R I E L K G I

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 2

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TACAACAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCTCTCGCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTTCTTGCCGTAGTT 1100

eGFPc.e.unc53sma

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGTCTCTGCGTCGCACGTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCAGCAGCAG 1200

eGFPc.e.unc53sma

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAAGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGGCGCCCT 1300

eGFPc.e.unc53sma

P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G

TCACCTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTT
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGCCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCCTAACCCGGTTAGCCGTGGAAAG 1400

eGFPc.e.unc53sma

C.e.unc53 sma

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D V A N R H L S

GAAGGGCAGCTTATCAAGTCGATTAGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500

eGFPc.e.unc53sma

C.e.unc53 sma

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCCTGCATTACGAAACGTTTGGCAAAAATCACATCGAACC TGGATGGCCTCGAAACGTGCTCGACTACCTGAAAAATC TGGGTC TCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTGTAGTGTAGCTTGGACCTACCGGAGCTTGCACAGAGCTGATGGACTTTTGTAGCCAGAGCTGACGA 1600

eGFPc.e.unc53sma

C.e.unc53 sma

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGTGCTCTTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTGCGCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCTGCTTTCGAAGCCGTTGACTT 1700

eGFPc.e.unc53sma

C.e.unc53 sma

S K L T K T D I D S S N L G A V L Q L L F L L S T Y K Q K L Q Q L I

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 1

AAAAGATCAGAAGAAATGGAGCAACTACCCACATCCATTATGCCACCCGGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTGAGCAACCGCTTCA
TTTTCTAGTCTTCTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT 180

eGFPc.e.unc53sma

C.e.unc53 sma

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCACATTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 190

eGFPc.e.unc53sma

C.e.unc53 sma

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAATAATACAAATTCATTCCGTCGCGTCGAGCGGTTCGAGTGGAATAATAA
TCGGTTTCTCGAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT 200

eGFPc.e.unc53sma

C.e.unc53 sma

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N M

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTGGCTGGATGGAGGGTTGAGGTTTTT 210

eGFPc.e.unc53sma

C.e.unc53 sma

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I

CCTTC TAGACCACAAACCCAGCTAGTTCTGTGTGCTACAAC TACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAACTTG
GGAAGATCTGGTGTGTTGGGTGATCAAGCACAACGATGTGATGTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC 220

eGFPc.e.unc53sma

C.e.unc53 sma

P S R P Q T O L V R V A T T T K I G S S K L A A P K A V S T P K L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCGATAACAGCGGTGGTGGTGGTGGGAATGCTGAAATTAAAGTTATTAGTAGCAAAAACCCATC
GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATTGTCGCCACCACCACCACCCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG 230

eGFPc.e.unc53sma

C.e.unc53 sma

A S V K T I G A K O E P D N S G G G G G G M L K L K L F S S K N P S

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 4

TTCTCATCGAATAGCCACAACCTACGAGAAAGGCGGCGGCTCAACAACAACTTTGTGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AAGGAGTAGCTTATCGGGTGTGGATGCTTTCCGCCGCCGCCACGGAGTTGTTGTTTGAACAGCTTTTAGCGACGGGGTCACTTTTCACGGACTTC 2400

eGFPc.e.unc53sma

C.e.unc53 sma

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L I

CCGCCGACCAGTAAGCTGGGAAGTGCCACGCTCTATGTGGAAGCTTTGTACGCCAAAAGTTTCTACCGTAAAACGGACGCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTGACCCCTTACGGTGCAGATACAGCTTCGAAACATGCGGTTTCAAAGGATGGCATTGCTGCGGGGTTAGTATAGAGTTGTT 2500

eGFPc.e.unc53sma

C.e.unc53 sma

P P T S K L G S A T S M S K L C T P K V S Y R K I D A P I I S Q Q

ACTCGAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTCTCAGGCCTATGCGACCTAAGTTGTCGTGACGCGTTGCAGTAGTAGCTGCCCTCCAAAGGATTCGTA 2600

eGFPc.e.unc53sma

C.e.unc53 sma

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCTCCATCGTGACAGCTATCAGACAGCGG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTGATAGTCTGTCGGC 2700

eGFPc.e.unc53sma

C.e.unc53 sma

F S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTGCGACAACCTCTTTTGGTTGTGACCGTCACCTTCTCACTTTTCGTGTCGCTTTTTTCTAG 2800

eGFPc.e.unc53sma

C.e.unc53 sma

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTGACAAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGATTCTTCAACTGTTTACTGGGGCACTATAGACT 2900

eGFPc.e.unc53sma

C.e.unc53 sma

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1>6960) Site and Sequence

Page 6

AAAACCAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTGGTCTTGGACTTTTCGAGGTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT 300

eGFPc.a.unc53sma

C.e.unc53 sma

K P E P E K L Q S M S I D T T D V P P L P P L K S V V P L K M T S

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCGTCTGAAGACT
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTGTAGTGTAGCGGACAGTTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA 310

eGFPc.e.unc53sma

C.e.unc53 sma

I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC
GGTAACACCGAGTACGCGCCGAGTCCACTGAGGCGGCTGTTTTGAAGACCATTAGTAAGCGACCTCTCTTCTACCCTTTCTTATTCGTAGCTTTAG 320

eGFPc.e.unc53sma

C.e.unc53 sma

S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S

CAGCGGCTACACCTCTGACGCCGGTGTTCGATGTGCGCCAAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTACTCCCTCTTCGACTTCTTATGCTACTGTACTGAGCAGCTCGTGCTTGGCCGATA 330

eGFPc.e.unc53sma

C.e.unc53 sma

S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A O N G Y

CCTGACAACCTCGAAGACAGTTCTCTCTGTCGCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCTGCTAAACAGGCCCTCATCTGTACC 340

eGFPc.e.unc53sma

C.e.unc53 sma

P D N F E D S S S L S S G I S D N N E L D D I S T D D L S G V D M

CAACAGTCGCTCCAACATAGCGACTATTTCCACTTTGTTGCGCATCCACGCTCTTCTTCTCAAAGCCCCGAGTCCCCAGTCGGTCTCCACATCACT
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA 350

eGFPc.e.unc53sma

C.e.unc53 sma

A T V A S K H S D Y S H F V R H P T S S S S K P R V P S R S S T S V

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1>6960) Site and Sequence

Page 6

CGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAACTTCTGTCCAGTGCCGAACGAGCCAACGTGGCGCGCTGCCACCTCAACCTTCGGACAA
GCTAAGAGCTAGAGCTCGTCTGTCTCTTACACATGTTGAAGACAGGGTCACGGCTTGCTCGGTTGCACGCGGCGACGGTGGAGTTGGAAGCCTGTT
3800

-----a3FPC.e.unc53sma-----
-----C.e.unc53 sma-----
D S R S R A E Q E N V Y K L L S Q C R T S Q R G A A A T S T F G Q
CATTTCGTAAGATCCCGGGATCCACGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCAC
GTAAGCGATTCTAGGGGCCCTAGGTGGCTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAACATCTCCAAATGAACGAAATTTTTTGGAGGGTG
3700

-----aGFPC.e.unc53sma-----
-----C.e.unc53 sma-----
H S L R S P G S T G S R . L I I I S H T T F V E V L L A L K N L P
ACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAGCAATAGCATCACAAA
TGGAGGGGACTTGACCTTTGATTTTACTTACGTTAACAACAACAATGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTT
3800
H L P L N L K H K M N A I V V V N L F I A A Y N G Y K . S N S I T N
TTTCACAAATAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATATTTGTT
3900
AAAGTGTATTATTCGTAAGGAGTACGTAAGATCAACACCAACAGGTTTGTAGTAGTTACATAGAATTGCGCATTTAACATTTCGCAATTATAAACAA
F T N K A F F S L H S S C G L S K L I N V S . R V N C K R . Y F V
AAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGCCGGAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGG
4000
TTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCCGTTTATAGGAATATTTAGTTTCTTATCTGGCTCTATCCC
K I R V K F L L N Q L I F . P I G R N R O N P L . I K R I D R D R
TTGAGTGTGTTCAGTTTGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTAC
4100
AATCACAACAAGGTCAAACCTGTTCTCAGGTGATAATTTCTGCACCTGAGGTTCAGTTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATG
V E C C S S L E O E S T I K E R G L O R Q R A K N R L S G R W P T T
GTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTTAAAGGGAGCCCCGATTAGAGCTTGACGGGGAAA
4200
CACTTGGTAGTGGGATTAGTTCAAAAACCCAGCTCCACGGCATTTCGTGATTAGCCTTGGGATTTCCTCGGGGGCTAAATCTCGAATGCCCCCTT
T I T L I K F F G V E V P . S T K S E P . R E P P I . S L T G K
GCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAACACACCC
4300
CGGCCGCTTGACCGCTCTTTCTTCCCTTCTTTCGCTTTCTCGCCCGCATCCGCGACCGTTACATCGCCAGTGCAGCGCATTGGTGGTGGG
A G E R G E K G R E E S E R S G R . G A G K C S G H A A R N H H T
GCCGCGCTTAATGCGCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTTGTTTATTTTCTAAATACATTCAATA
4400
CGGCGCAATTACCGCGCGATGTCGCGCAGTCCACCGTGAAGAGCCCTTTACACGCGCTTGGGGATAAACAAATAAAAGATTTATGTAAGTTTAT
R R A . C A A T G R V R V H F S G K C A R N P Y L F I F L N T F K Y
TGATCCGCTCATGAGACAATAACCTTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTGAGGCGGAAAGAACAGCTGTGGAATGTGTGTCAG
4500
ACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTCTCTCAGGACTCCGCTTCTTGGTCGACACCTTACACAGATC
V S A H E T I T L I N A S I I L K K E E S . G G K N Q L W N V C Q

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 7

TTAGGGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAAGGTGTGGAAGTCCCCAGGCTCCCC
AATCCCACACCTTTCAGGGGTCGAGGGGTCGTCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCGAGGG
L G C G K S P G S P A G R S M Q S M H L N . S A T R C G K S P G S 480

CAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCGCCCTAACGCCCATCCGCCCTAACGCCCCAGTTCGCG
GTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGG
P A G R S M Q S M H L N . S A T I V P P L T P P I P P L T P P S S A 470

CCATTCTCGGCCCATGGCTGACTAATTTTTTTTATTTATGAGAGGCGGAGGCCCTCGGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT
GGTAAGAGGCGGGGTACCGACTGATTAATAAAAAATAAATACGCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCCTCCGAAAAA
H S P P H G . L I F F I Y A E A E A A S A S E L F Q K . . G G F F 460

GGAGGCTTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCGCCGCCGCTTG
CCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTTAGCAAAGCGTACTAATTGTTCTACCTAACGTGCGTCCAAGAGGCGGGGAACC
G G L G F C K D R S R D R M R I V S H D . T R W I A R R F S G R L 450

GTGGAGAGGCTATTGGCTATGACTGGGCACAACAGACAATCGGCTGCTGTATGCCGCCGTGTCCGGCTGTGAGCGCAGGGGCGGCCGTTCTTTTGG
CACCTCTCCGATAAGCCGATCTGACCCGTTGTCTGTGTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCGCGGGCCAAGAAAAAC
G G E A I R L . L G T T D N R L L . C R R V P A V S A G A P G S F C 500

TCAAGACCGACC TGTCGGTGCCCTGAATGAAGTGAAGACGAGGCAGCGCGGTATCGTGGCTGGCCACGACGGCGTTCTTGCGCAGCTGTGCTCGA
AGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCT
O D R P V R C P E . T A R R G S A A I V A G H D G R S L R S C A R 510

CGTTGTCACTGAAGCGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCTGTCACTCACCTTGCTCTCGCGAGAAAGTATCCATC
GCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGCCCCGTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAG
R C H . S G K G L A A I G R S A G A G S P V I S P C S C R E S I H 520

ATGGCTGATGCAATGCGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG
TACCGACTACGTTACGCCCGGACGATGCGAAGTACGGCGATGGACGGGTAAAGTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACC
H G . C N A A A A Y A . S G Y L P I R P P S E T S H R A S T Y S D G 530

AAGCGGTCTTGTGCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGAGCATGCCGACGGCGA
TTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTCGTAGTCCCGAGCGGGTCCGCTTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCGGCT
S R S C R S G . S G R R A S G A R A S R T V R Q A Q G E H A R R R 540

GGATCTCGTCGTGACCCATGGCGATGCTGCTTGGCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCG
CCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGCGAAAAGACCTAAGTAGCTGACACCGGCGACCCACACCGC
G S R R D P V R C L L A E Y H G G K W P L F V I H R L V P A G C G 550

GACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGCTC
CTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAAGCGCCGCTTACCCGACTGGCGAAGGAGCAGCAATGCCATAGCGGCGAG
G P L S G H S V G Y P . Y C . R A V R R M G . P L P R A L R Y R R S 560

CCGATTGCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCA
GGCTAAGCGTCGCTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTTCGTCGCGGTTGGACGGT
R F A A H R L L S P S . R V L L S G T L G F E M T D Q A T P N L P 570

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 4

TCACGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCA
AGTGCTCTAAAGCTAAGGTGGCGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGT 580
S R D F D S T A A F Y E R L G F G I V F R D A G W M I L Q R G D L
TGCTGGAGTTCTTCGCCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAGACAGAA
ACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCTTCTTGGGCGGATACTGCCGTTATTTTCTGTCTT 590
M L E F F A H P R G R L T E T R K E T I P E G T R A M T A I K R Q H
TAAACGCACGGTGTGGGTCGTTTGTTCATAAACCGGGGTTCCGTTCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCATTGGGGCCAATAC
ATTTTGCCTGCCACAACCCAGCAACAAGTATTGCGCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGTTATG 600
K T H G V G S F V H K R G V R S Q G V H S V D T P P R P H W G Q Y
GCCCGCGTTTCTTCTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACGTGGGGCGGCAGGCCCTGCCATAGCCTCAG
CGGGCGCAAGAAGGAAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTCGGTTGCAAGCCCGCGCTCCGGGACGGTATCGGAGTC 610
A R V S S F S P P H P P S S G E G P G L A A N V G A A G P A I A S
GTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTA
CAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAACTATTAGAGTACTGGTTTATAGGAAT 620
G Y S Y I L I D L K L H F F K R I V K I L F D N L M T K I P
ACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGAACA
TGCATCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTTAGAAGAACTCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTGT 630
R E F S F H A S D P V E K I K G S S D P F F L R V I C C L Q T
AAAAAACCCCGCTACCAGCGGTGGTTTGTTCGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGGCTTCAGCAGAGCGCAGATACCAATA
TTTTTGGTGGCGATGGTCCCAACCAACACGGCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGCTCTCGCTCTATGGTTAT 640
K K P P L P A V V C L P D Q E L P T L F P K V T G F S R A Q I P H
CTGCTCTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCTGCTGC
GACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGATGGAGCGAGACGATTAGGACAATGGTCACCGACGACG 650
T V L L V P L G H H F K N S V A P P T Y L A L L I L L P V A A A
CAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGTTCTGTCACACAGCC
GTCACCGCTATTGACACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCTATTCCGCTGCGCAGCCGACTTGCCCCCAAGCAGGTGTGTGGG 660
S G D K S C L T G L D S R R L P D K A Q R S G T G G S C T Q P
AGCTTGGAGCGAACGACCTACACGAAC TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGCGGACAGGTATCCGG 670
TCGAACCTCGCTTGTGATGTGGCTTGACTCTATGGATGTCGACCTGATACCTTTTCGGGTGCGAAGGGCTTCCCTCTTCCGCTGTCCATAAGCC
S L E R T T Y T E L R Y L Q R E L E S A T L P E G R K A D R Y P
TAAGCGGCAGGGTCGGAACAGGAGAGCGCAGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCCGTGCGGGTTTCGCCACCTCTGACTTGA 680
ATTCGCCGTCACGCTTGTCTCTCGCTGCTCCCTCGAAGGTCCCCCTTTCGGGACCATAGAAATATCAGGACAGCCAAAGCGGTGGAGACTGAGCT
V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L L E
GCSTCGATTTTTGTGATGCTCGTCAAGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCCTGGCTTTTGTGCGCTTTTGT 690
CGCAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTTCGGTCTGTCGCGGAAAAATGCCAAGGACCGGAAAAACGCGGAAAAACGA
R R F L C S S G G R S L V K N A S N A A F L R F L A F C W P F A

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 6

• CACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
GTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 6960
H M F F P A L S P D S V D N R I T A M H

Tuesday, 18 November 1997 11:46

fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Enzymes: 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

3p.

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGCGC 100
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGC TGGGGGCGGGTAAC TGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAAC TGCAGTTACCCACCTCATAAATGCCA 200
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACTGTCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCAT 300
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCAC TACGCCAAAACCGTCATGTAGTTACCCGACCT 400
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W
TAGCGGTTTGACTCACGGGGATTTC AAGTCTCCACCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGCTA
ATCGCCAAATGAGTGCCCTAAAGGTT CAGAGGTGGGTAAC TGCAGTTACCTCAAAACAAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT 500
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S
ACAAC TCCGCCCCATTGACGCAATGGGCGGTAGGCGGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAAC TCGGTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTCTGCTCGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT 600
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
CCSGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGAGCTAAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCTCGCTCTCGACAAGTGGCCCCACACGGGTAGGACCACTCGACCTGCCGCTGCATTGCGCGGTGTTCAAGTCGC 700
eGFPc.e.unc53ec1
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACGGCAAGCTGCCCGTGGCCACCTCGTGAC
ACAGGCGCTCCCGCTCCCGCTACGGTGGATGCCGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCGGTCGACGGGACGGGACCGGTTGGGAGCACTE 800
eGFPc.e.unc53ec1
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
CACCTTGACCTACGGCGTGCAGTGCTTACGCCGTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAG
GTGGGAC TGGATGCCGACGTACGAAGTCGGCGATGGGGCTGGTGACTTCGTCTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCTCTC 900
eGFPc.e.unc53ec1
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A P P E G Y V Q E
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGSGCATCG
GCGTGGTGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGGCTAGCTCGACTTCCCGTAGC 1000
eGFPc.e.unc53ec1
R T I F F K D D G N Y K T R A E V K F E G D T L V H R I E L K G I

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 1

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCA
TGAAGTTCTCTCGCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTGTCAGATATAGTACCGGCTGTTCTGCTCTCTTGCCTAGT

110

eGFPC.e.unc53ec1

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCTGCGCTGCACGTCGAGCGGCTGGTGATGGTCTGTTGGGGGTAGCCGCTGCGGGGACGACGAC

120

eGFPC.e.unc53ec1

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTCC TGCTGGAGTTCGTGACCGCGCGGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGATACAGGACGACCTCAAGCACTGGCGGGCGGCT

130

eGFPC.e.unc53ec1

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCTTAACCCGGTTAGCCGTGGAAAG

140

eGFPC.e.unc53ec1

C.e.unc53 ec1

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTACGCTAATCCCTATAAAGGTTACTAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCT

150

eGFPC.e.unc53ec1

C.e.unc53 ec1

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGECTGCATTACGAAACGTTTGCAAAAAATCACATCGAACC TGGATGGCTCGAAACGTGTC TCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAAACGTTTTAGTGATAGCTTGGACCTACCGGAGCTTTCACAGAGCTGATGGACTTTTACACCCAGAGCTGACGA

160

eGFPC.e.unc53ec1

C.e.unc53 ec1

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTCTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTGCGCTTTGAACCCACGTCAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCTGCTTCGAAGCCGTTGACT

170

eGFPC.e.unc53ec1

C.e.unc53 ec1

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K O K L R Q L I

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 8700) Site and Sequence

Page 3

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCTCGCCACGTGTCGCCACGTCAAGCAACCGTTCA
TTTTC TAGTCTTCTTTAACCTCGTTGATGGGTGATAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTGGCGAAGT 1800

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACCTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA
CGTTGAT TGGGTTTAAGGTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 1900

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAATAATACAAATTCATTCCGTCGTCGAGCCGTTGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTACCGTTATTAT 2000

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTGGAATCTAAACCGACCTACCTCCCAACTCCAAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTGGCTGGATGGAGGTTGAGGTTTTT 2100

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I

ECTTCTAGACCACAAACCCAGCTAGTTTCGTGTTGCTACAAC TACAAAAATCGGAAGCTCAAAGCTAGCCGC TCCGAAAGCCGTGAGCACCCCAAACTTG
GGAAGATCTGGTGTGTTGGGTCGATCAAGCACAACGATGTTGATGTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGTTTTGAA 2200

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

P S R P Q T Q L V R V A T T T K I G S S K L A A P K A V S T P K L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGAATGCTGAAATTAAGTTATTAGTAGCAAAAACCATC
GAAGACACTTCTGATAACCTCGTTTTGTTCTCGGGCTATGTGCGCCACCACCACCACCTTACGACTTAATTTCAATAAGTCATCGTTTTTGGGTAG 2300

—eGFPC.e.unc53ec1—

—C.e.unc53 ed—

A S V K T I G A K Q E P D N S G G G G G G M L K L K L F S S K N P S

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 9

TTCTCATCGAATAGCCACAACTACGAGAAAGGCGGCGGCGGTGCCTCAACAACAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AASGAGTAGCTTATCGGGTGTGGATGCTCTTTCGCCGCCGCCACGGAGTTGTTGTTTGAACAGCTTTTAGCGACGGGGTCACTTTCACCGGACTTC
240
eGFPC.e.unc53edl
C.e.unc53 edl
S S S N S P O P T R K A A A V P Q Q Q T L S K I A A P V K S G L K
CCGCCGACCAGTAAGCTGGGAAGTGCCACGTCATGTGCAAGCTTTGTACGCCAAAAGTTCTACCGTAAAACGGACGCCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTGACCCCTTACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTCCTGCGGGGTTAGTATAGAGTTGTTCT
250
eGFPC.e.unc53edl
C.e.unc53 edl
P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q
ACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTCTCAGGCCATGCGACCTAAGTTGTCGTGCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTTCGTA
260
eGFPC.e.unc53edl
C.e.unc53 edl
D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M
GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCCTCCATCGTGACAGCTATCAGACAGCGG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTGATAGTCTGTCGGC
270
eGFPC.e.unc53edl
C.e.unc53 edl
H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P
ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTTCGGACAACCTCTTTTGGTTGTGACCGTCACCTTCTCACTTTTCGTGTCGCTTTTCTAG
280
eGFPC.e.unc53edl
C.e.unc53 edl
I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D
CAGCTCCAGCTGTTCCGCCAGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTGACAAAGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGATTCTTCAACTGTTTACTGGGGCACTATAGACT
290
eGFPC.e.unc53edl
C.e.unc53 edl
P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

AAAACCGAACC TGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAAATCAGTTGTTCCACTTAAAAATGACTTCA
TTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTACTGAAGT 3000
eGFPC.e.unc53ec1
C.e.unc53 ed
K P E P E K L Q S M S I D T T D V P P L P P L K S V V P L K M T S
ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAGGAAAAATCACATCGCCTGTCAAGTCGTTGGATATGAGCAGTCGTCGCGTCTGAAGACT 3100
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTCTCTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGCGCAGACTTCTGA
eGFPC.e.unc53ec1
C.e.unc53 ed
I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D
CCATTGTGGCTCATCGCTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTCTGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC 3200
GGTAACACCGAGTACGCAGCGGAGTCCACTGAGGCGGCTGTTTTGAAGACCATAGTAAGCGACCTCTCTCTACCCTTCTTATTCTGTAGTCTTAG
eGFPC.e.unc53ec1
C.e.unc53 ed
S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S
CAGCGGCTACACCTCTGACGCCGGTGTGCGATGTGCCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT 3300
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTACTCCCTCTTCGACTTCTTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA
eGFPC.e.unc53ec1
C.e.unc53 ed
S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A Q N G Y
CCTGACAACCTCGAAGACAGTTCCTCTTGTGCTGCTGGAATATCCGATAACAACGAGGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATAEC 3400
GGACTGTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCCCTTAGGTGGCTTAGATCTATTGACTAGTATTAGTCGGTATGG
eGFPC.e.unc53ec1
C.e.unc53 ed
P D N F E D S S S L S S G I S D N N E S I H R I . I T D H N Q P Y
ACATTGTAGAGGTTTTACTTGTCTTAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATG 3500
TGTAACATCTCCAAATGAACGAAATTTTGGAGGGTGTGGAGGGGACITGGACTTTTATTTTACTTACGTTAACAACAACAAATGAACAAATAAC
H I C R G F T C F K K P P T P P P E P E T . N E C N C C C . L V Y C
CAGCTTATAATGGTTACAAATAAGCAATAGCATCACAATTTTCAAAATAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA 3600
GTGGAATATTACCAATGTTTATTCGTTATCGTAGTGTAAAGTGTTTATTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGTAGTAGT
S L . W L Q I K O . H H K F H K . S I F F T A F . L W F V O T H Q

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 6

TGTATCTTAACGCGTAAATGTAAGCGTTAATATTTTGTAAAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGG
ACATAGAATTGCGCATTTAACATTGCAATTATAAAACAATTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCC 370
C I L T R K L . A L I F C . N S R . I F V K S A H F L T N R P K S
CAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCAGTTTGAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTC
GTTTATAGGAATATTTAGTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTCTCAGGTGATAATTTCTGCACCTGAGGTGCA 380
A K S L I N Q K N R P R . G . V L F Q F G T R V H Y . R T W T P T S
AAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGGGTTCGAGGTGCCGTAAAGCACTAAATCGGA
TTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGTGATTAGCCT 390
K G E K P S I R A M A H Y V N H H P N Q V F W G R G A V K H . I G
ACCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGTAGGGCGCT
TGGGATTTCCTCGGGGGCTAAATCTCGAACTGCCCTTTCGGCCGCTTGACCGCTCTTCTCTCCCTTCTTCGCTTTCCTCGCCCGCATCCCGCGA 400
T L K G A P D L E L D G E S R R T V R E R K G R K R K E R A L G R
GGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGC
CCGTTACATCGCCAGTGCAGCGCATTGGTGGTGTGGGCGGCGCAATTACGCGCGCATGTCCCGCGCAGTCCACCGTGAAAAGCCCCCTTACACGCG 410
V Q V . R S R C A . P P H P P R L M R R Y R A R Q V A L F G E M C A
GGAACCCCTATTTGTTATTTTCTAAATACATTCAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGA
CCTTGGGGATAAACAATAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTTCTTCT 420
E P L F V Y F S K Y I Q I C I R S . D N N P D K C F N N I E K G R
GTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCGAGGCTCCCGAGCAGGCGAGAAGTATGAAAGCATGCATCTC
CAGGACTCCGCTTCTTGGTTCGACACCTTACACACAGTCAATCCACACCTTTAGGGGTCCGAGGGGTCGTCGCTTTCATACGTTTCGTACGTAGAG 430
V L R R K E P A V E C V S V R V W K V P R L P S R Q K Y A K H A S
AATTAGTCAGCAACCAGGTGTGGAAGTCCCGAGGCTCCCGAGCAGGCGAGAAGTATGAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TTAATCAGTCGTTGGTCCACACCTTTAGGGGTCCGAGGGGTGTCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGG 440
Q L V S N Q V W K V P R L P S R Q K Y A K H A S Q L V S N H S P A P
TAATCCGCCCATCCCGCCCCTAATCCGCCCAGTTCCGCCCATTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTC
ATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAATAAATAAATACGCTCTCCGGCTCCGGCGGAG 450
N S A H P A P N S A Q F R P F S A P V L T N F F Y L C R G R G R L
GGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
CCGGAGACTCGATAAGGTCTTATCATCTCTCCGAAAAACCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTTAGCAAGCGTACTA 460
G L . A I P E V V R R L F V R P R L L O R S I K R Q D E D R F A
TGAACAAGATGGATTGCACGACGTTCTCCGGCCGCTTGGTGGAGAGGCATTCCGGCTATGACTGGGCAACAGACAATCGGCTGCTCTGATGCCGCC
ACTTGTCTACCTAACGTGCGTCCAAGAGGCCGCGCAACCCACCTCTCCGATAAGCCGATACGACCGGTGTTGCTGTTAGCCGACGAGACTAGGGCG 470
L N K M D C T O V L R P L G V R G Y S A M T G H N R O S A A L M P P
GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAAC TGCAAGACGAGGCGAGCGCGCTATCGT
CACAAGGCCGACAGTCGCTCCCGCGGGGCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGCTCGCGCCGATAGCA 480
C S G C O R R G A R F F L S R P T C P V P . M N C K T R O R G Y R

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 7

GGCTGGCCACGACGGGCTTCCTTGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
CCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGATAACCCGCTTCACGGCCCCGCTCTAGA
G W P R R A F L A Q L C S T L S L K R E G T G C Y V A K C R G R I 490

CCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCATTTCGACCAC
GGACAGTAGAGTGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGATGCGAACTAGGCCGATGGACGGGTAAGCTGGTG
S C H L T L L L P R K Y P S V L M Q C G G C I R L I R L P A H S T T 500

CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCG
GTTGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGGGTCTGGC
K R N I A S S E H V L G W K P V L S I R M I V T K S I R G S R Q P 510

AACTGTTCCGACGGCTCAAGGCGAGCATGCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAATGGCCG
TTGACAAGCGGTCCGAGTTCGCTCGTACGGGCTGCCGCTCTTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGG
N C S P G S R R A C P T A R I S S . P H A M P A C R I S W V K M A 520

CTTTTCTGGATTATCGAGCTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGCGGGCGAA
GAAAAGACCTAAGTAGCTGACACCGGCGGACCCACACCGCTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCGGCTT
A F L D S S T V A G W V W R T A I R T . R V L P V I L L K S L A A N 530

TGGGTGACCGCTTCCTCGTCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGCGGAGGGCTAAGCGTCTCGTAGCGGAAGATAGCGGAAGAAGTGTCAAGAAGACTTCGCCCTGAGA
G L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S 540

GGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCACGAGATTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTT
CCCCAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAG
G V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F 550

CGGGACCGCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCCCTGCGGGCGACCTACTAGGAGGTGCGGCCCTTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCGGATTGACTTTGTGCTTCTCTGTTATG
P G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T 560

CGGAAGGAACCCGCGCTATGACGGCAATAAAAGACAGATAAAACGCACGGTGTGGGTGCTTTGTTCATAAACGCGGGGTTGGTCCCAGGGCTGGGA
GCC TTCTTGGGCGCGATACTGCCGTTATTTTCTGTCTTATTTTGGCTGCCACAACCCAGCAACAAGTATTGCGCCCCAAGCCAGGGTCCCAGCCGT
G R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A 570

CTCTGTCGATACCCACCGAGACCCATTGGGGCCAATACGCCGCGTTCCTTCTTTTCCCACCCACCCCAAGTTGGGTGAAGGCCAGGGCTC
GAGACAGCTATGGGGTGGCTCTGGGGTAACCCGGTTATGCGGGCGCAAGAAGGAAAGGGGTGGGGTGGGGGTTCAAGCCCACTTCGGGTCCCAGS
L C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A 580

GCAGCCAACGTCGGGGCGGAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTAAACATTCATTTTAAATTTAAAGGATCTAGS
CGTCGGTTCGAGCCCCCGCTCCGGGACGGTATCGGAGTCCAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCC
R S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G 590

TGAAGATCTTTTGTATAATCTCATGACCAAAATCCCTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTCAGAAAAGATCAAGGATCTTCTT
ACTTCTAGGAAAACTATTAGAGTACTGGTTTATGGGAATTGCACTCAAAAGCAAGGTGACTCGAGTC TGGGGCATCTTTCTAGTTTCTTAGAAGAAC
E D P F . . S H D O N P L T . V F V P L S V R P R R K D Q R I F L 600

Tuesday, 18 November 1997 11:46
fig 32 pEGFPect (1 > 6700) Site and Sequence

Page 8

AGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCCTACCAGCGGTGGTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTT
TCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTTGTGTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 8100
R S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F

CCGAAGGTAAGTGGCTTACGAGAGCGCAGATACCAATACTGCTCTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAGAACTCTGTAGCACCCTA
GGCTTCCATTGACCGAAGTCGCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTCTTGAGACATCGTGGCGGAT 8200
S E G N W L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y

CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTGGACTCAAGACGATAGTTACCGGATAAGGC
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTACGACAGAAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCG 8300
I P R S A N P V T S G C C Q V R . V V S Y R V G L K T I V T G . G

GCAGCGGTGCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAG
CGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTCGGGTGCAACCTCGCTTGTGATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCC 8400
A A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K

GCCACGCTTCCCGAAGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGTGGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCTGGT
CGGTGCGAAGGGCTTCCCTCTTTCCGCTGTCCATAGGCCATTCCGCGTCCAGCCTTGCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCA 8500
R H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V

ATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGC
TAGAAATATCAGGACAGCCCAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTTGCGGTGCTTGGC 8600
S L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R

GGCCTTTTACGGTTCTTGGCCTTTTGTGCGCTTTTGTACATGTTCTTTCCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CCGGAAAAATGCCAAGGACCGGAAACGACCGGAAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 8700
G L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence
Enzymes : 72 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

TAGTTATTAAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTTCATGCGGGGATAACGTCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W
TAGCGGTTTGACTCACGGGGATTTCAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCTAAAGGTTTCAGAGGTGGGGTAACGTCAGTTACCTCAAACAAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .
ACAAC TCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGCTATATAAGCAGAGCTGGTTTGTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTCGTCGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT
Q L R P I D A N G R . A C T V G G L Y K Q S W F S E P S D P L A L
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGACGTAACGGGCCACAAGTTCAGCG
GGCAGCGGTGGTACCACTCGTTCCCGCTCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTGGCCGGTGTCAAGTCCG
eGFPC.e.unc53xba
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCCGGCGAGGGCGAGSGCGATGCCACCTACGGCAAGCTGACCTGAAGTTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCTCGTGAC
ACAGGCGGCTCCCGCTCCCGCTACGGTGGATGCCGTTCGACTGGGACTTCAAGTAGACGTGGTGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG
eGFPC.e.unc53xba
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T
CACCTGACCTACGGCGTGCAGTGCTTCAGCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCTGTCTGAAGAAGTTCAGGCGGTACGGGCTTCGATGCAGGTCTCT
eGFPC.e.unc53xba
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTGAAGAAGTTCCTGCTGCCGTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC
eGFPC.e.unc53xba
P T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 1

ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGCTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTCTTCCGCTAGTT 110

eGFPC.e.unc53xba

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCACGGCCCGTGTCTGCTG
CCACTTGAAGTTCTAGCGGTGTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGACGACGAC 120

eGFPC.e.unc53xba

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTACGGCGGGAAGCTGTTCTGGGGTGTCTTTCGCGCTAGTGATACAGGACGACCTCAAGCACTGGCGGGCGGCCCT 130

eGFPC.e.unc53xba

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTAGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCTAACCCTGTTAGCCGTGGAAAG 140

eGFPC.e.unc53xba

C.e.unc53 xba

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTACGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 150

eGFPC.e.unc53xba

C.e.unc53 xba

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCGCTGCAATTCACGAAACGTTTGGCAAAATCACATCGAACCTGGATGGCTCGAAACGTGCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACGTTTGTAGTGAGCTTGGACCTACCGGAGCTTGCACAGAGCTGATGGACTTTTAGACCCAGAGCTGACGA 160

eGFPC.e.unc53xba

C.e.unc53 xba

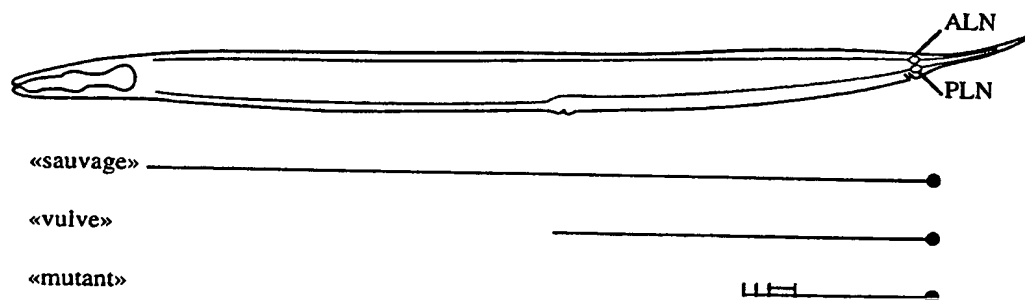
F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGTGCTCTTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTGGCTATAGCTGTGCTTTGAACCCACGTCAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTGCTCTTGAAGCCGTGACTT 170

eGFPC.e.unc53xba

C.e.unc53 xba

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L I



souches	phénotypes	sauvage	vulve	mutant	nombre
wt	ALN	100	0	0	70
	PLN	100	0	0	70
<i>unc-53(n152)</i>	ALN	0	26	74	50
	PLN	nd	nd	nd	
<i>unc-53(n152)</i> <i>pAΔunc-53</i>	ALN	95	0	5	65
	PLN	77	0	23	44
<i>unc-53(n152)</i> <i>pAΔunc-53-H1</i>	ALN	25	60	15	56
	PLN	52	16	32	25

Figure 51b

Tuesday, 18 November 1997 13:56

fig 52 pLM5 (1 > 5425) Site and Sequence

Enzymes : All 146 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

GACGGATCGGGAGATCTCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCC TAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGCAGGAGCAAAATTTAAGCTACAACAGGCAAGGCTTGACCGACAATGTCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCAGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAAC TGCTGTTAAGCTACTTCTTAGACGAATCCCAATCCGCAAAACGC
G G R . V V R E Q N L S Y N K A R L D R Q L H E E S A . G . A F C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGTCTATATGCCGAACGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A W L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACGTGACGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACGTCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTGATGTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTTGACTCACGGGGATTTCAGTCTCCACCCCATTTGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGACCTATCGCCAAACTGAGTGCCCTAAAGGTTACAGAGTGGGGTAACGTCAGTT
R Y Y H G D A V L A V H Q W A V I A V . L T G I S K S P P H . R Q

TGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCGCCCCATTGACGCAATGGGCGGTAGGCGGTACGGTGGGAG
ACCTCAAAACAAAACCGTGGTTTTAGTTGCCCCGAAAGGTTTTACAGCATGTTGAGGCGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCTC
V E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGG
CAGATATATTCGCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGACCGGATCG
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S W L A

GTTTAAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGATGGTATGGCTAGCATGACTGGTGGACAGCAATGGGTGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCC TAGACATGCTG
F K L K L T M G G S H H H H H H G M A S M T G G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 2

GATGACGATAAGGTACCTAGGATCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGGCTCTGAGCTATGGGAGAAGGAAATGA
CTACTGCTATTCCATGGATCCTAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCTTTAC 100
pLM5 insert = U3
D D D K V P R I H A N E E E E P E K K E V S E L R S E L V E K E M
AGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGA4
TCGAATGCTGTAGGCGAACCCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACT 120
pLM5 insert = U3
ORF U3
K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L E V D L L L
AGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCC
TCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACC TAGTAGACGTAATAGAAGGGGTGCGGCAGG 140
pLM5 insert = U3
ORF U3
A E N D R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S
CTAGGCCCTGGCACTCACCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGA
GATCCGGACCGTGAGTGGTAAGGAAGCCGGGGTCAGAAGCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCTCTTCACT 160
pLM5 insert = U3
ORF U3
L G L A L T H S F G P S L A D T D L S P M D G I S T C G P K E E V
CCCTCCGGGTGGTGGTGAGGATGCCCCSCAGCACATCATCAAGGGGACTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGCAAGGTGAGTGGAAAAGT
GGGAGGCCACCACCACTCTACGGGGCGTCGTGAGTAGTTTCCCCTGAACCTCGTCGCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCA 180
pLM5 insert = U3
ORF U3
T L R V V V R M P P Q H I K G D L K Q Q E F F L G C S K V S G F V
TGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATC
ACTGACCTTCTACGACCTACTTCGACAAAAGGTTTCAAGTTCTTGATATAAGATTTCCTGGGTCGGAGATGGGACCCCTGATTCGTGACTCAGGTAG 200
pLM5 insert = U3
ORF U3
D V K M L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 3

CATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCA
GTACCGATGTCGTAGTCGGTGCACCTTGTCTACAACCTACGCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAG 700

pLM5 insert = U3

ORF U3

H G Y S I S H V K R V L D A E P P E M P P C R R G V N N I S V S L

AAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCAGG 1200

TTCCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTGGCTACTACGTCGTGATGTATTGGAGGACGACTTCGTGGCCGC

pLM5 insert = U3

ORF U3

K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R

CCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGC 1300

GGAGCAGGAGAGCCCGGGTCGCCGTGCCCGTCTGGATGGACTGGTTAGCGAACCAGGCTCATGGACCACC TC GCGAGACCGGCACTCCAGTGTCTCCCG

pLM5 insert = U3

ORF U3

L V L S G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G

ATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGAATTGGGG 2000

TAGCAGTCGTGGAAGTGTACGTGGTCTGCAGAACGTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCC

pLM5 insert = U3

ORF U3

I V S T F N M H Q Q S C K D L Q L Y L S N L A N O I D R E T G I G

ATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTA 2100

TACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAAGTTACCCGGGAGTGGACGTTTCATAGTATTACAGGGA

pLM5 insert = U3

ORF U3

D V P L V I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y

TATTATAGGTACCAACATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCC 2200

ATAATATCCATGGTGGTGTAGTCGGACATTTTACTGTGGGTGGTACCAGACGTGAACTCGAAGTCC TACAAC TGGAGAGGTTGTTGCACCTCGGTCCG

pLM5 insert = U3

ORF U3

I I G T T N Q P V K M T P N H G L H L S F R M L T F S N N V E P A

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 4

AATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACGAAGAGCTGCTCGGGTGCCTCGACTGGGTAC
TTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTTCGACGAAGCCACGAGCTGACCCATG 220

pLM5 insert = U3

ORF U3

N G F L V R Y L R R K L V E S D S D I N A N K E E L L R V L D V V

CCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTCTGTCGTGCCATTGGCATTGA
GGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGAACGAAGAAAGACAGCACAGGTAACCGTAACCT 240

pLM5 insert = U3

ORF U3

P K L W Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E

GGACTTCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCT
CCTGAAGGCTTGGACCAAGTAACGGACACCTTGTGAGATAGTAAGGATAGATGTCTTCTCGGTTCTTACCCTATTTCAGGTACCTGTCTTTCGA 250

pLM5 insert = U3

ORF U3

D F R T V F I D L W N N S I I P Y L Q E G A K D G I K V H G Q K A

GCTTGGGAGGACCCAGTGGATGGGTCCGGGACACACTTCCC TGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCACCGTGG
CGAACCTCTCTGGGTACCTTACCCAGGCCCTGTGTGAAGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACC 260

pLM5 insert = U3

ORF U3

A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P P T V

GECCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAACCT
CGGGAGTGTCTGAACGGAGTGGAGGCTCTATCCTGTCAAGTTCTGTCTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGA 270

pLM5 insert = U3

ORF U3

G P H S I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L

TCAAGAAGCTGCCAAC TACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTACCCCT
AGTCTTCGACGGTGTAGTGAACCTCAGAGGTCTAGCTCTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATTCCTAAGCCGTTAGTGACAGTGGGG 280

pLM5 insert = U3

ORF U3

D E A A N Y I E S P D R E T I L D P N L O A T L . G F G N H C H P

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 5

CGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAG
GCC TGTCTCTTGCACCGTAGTCGATAGAAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGGGGTCTCTCTTGTCTCTCTCTC
pLM5 insert = U3
R T A E R V H Q L S . L L L S P L L F Q S T G S P A P G G E Q E G
GAGGAGATGAAGAGGAGGACAGGTCTTTGGTGCTGTACCTTTGAGAATTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTGTGCCCCCTAAA
CTCTCTACTTTCTCTCTCTCTGTCGAAGAACACGACATGGAACCTCTGAAGGATCTTCTTACCACCCACCACCAACCCCTTGAACACGGGGGATTT
pLM5 insert = U3
G G D E R G G T G S V C C T F E N F L G R N G G V A F G N L C P L N
CACATTTACTGGCTCTCTAGAGGGCCGTTTAAACCCGCTGATCAGCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCT
GTGTAAATGACCGGAGGAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAACGGGGAGGGGGCA
pLM5 insert = U3
T F T G L L . R A R L N P L I S L D C A F . L P A I C C L P L P R
GCCTTCTTGACCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
CGGAAGGAAC TGGGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTACTCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCC
A F L D P G R C H S H C P F L I K . G N C I A L S E . V S F Y S G
GGTGGGGTGGGGCAGGACASCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA TGGGTGGGCTCTATGGCTTC TGAAGCGGAAAGAACCA
CCACCCACCCCGTCTGTCTGTTCCCTTCTAACCCTTCTGTTATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCTTTCTTGST
G V G G A G Q Q G G G L G R Q . Q A C V G C G G L Y G F . G G K N Q
GCTGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGG
CGACCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGCGCGCCGACACCAATGCGCGTCGCATGSGGATGTGAACGGTGGG
L G L . G V S P R A L . R R I K R G G C G G Y A Q R D R Y T C Q R
CCTAGCGCCGCTCCTTTCTGCTTTCTTCCCTTCTTCTCGCCACGTTGCGCGGCTTTCCCGCTCAAGCTCTAAATCGGGGATCCCTTTAGGGTTCCGA
GGATCGCGGGCGAGGAAAGCGAAAGGAAGGAAGGAAGAGCGGTGCAAGCGGCCGAAAGGGGCAAGTTCGAGATTTAGCCCCGTAGGGAATCCCAGGGCT
P S A R S F R F L P F L S R H V R R L S P S S S K S G H P F R V P
TTTAGTGCTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCCCTTTGACGT
AAATACGAAATGCCGTGGAGCTGGGGTTTTTGAAC TAATCCAC TACCAAGTGCATCACC CGTAGCGGGACTATCTGCCAAAAAGCGGAAACTGGA
I . C F T A P R P O K T . L G . W F T . W A I A L I D G F S P F D V
TGGAGTCCACGTCTTTAATAGTGGACTCTTGTTCAAAACCTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGGGAT
ACCTCAGGTGCAAGAAATTATCAGCTGAGAACAAGGTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAATAAATATCCCTAAACCCCTA
G V H V L . . V T L V P N W N N T Q P Y L G L F F . F I R D F G D
TTGGGCTATTGGTTAAAAATGAGCTGATTTAACAAAAATTAACGCAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCAAGGT
AAGCCGGATAACCAATTTTTACTCGACTAAATGTTTTTAAATTCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTGAGGGTCCGA
F G L L V K K . A D L T K I . R E L I L W N V C O L G C G K S P G

Tuesday, 18 November 1997 13:58
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 6

CCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
GGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCTCCGTCTTCATACGTTTCGT 300
S P G R Q K Y A K H A S Q L V S N Q V W K V P R L P S R Q K Y A K H
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGCCCATGTCGCCCATTCGCCCATGGCTGACT
ACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGA 400
A S Q L V S N H S P A P N S A H P A P N S A Q F R P F S A P V L T
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCCTAGGCTTTTGCAAAAAG
TTAAAAAAATAAATACGTTCTCCGGCTCCGGCGGAGACGAGACTCGATAAGGTCTTCATCACTCTCCGAAAAACCTCCGGATCCGAAAACTGTTTTT 500
N F F Y L C R G R G R L C L . A I P E V V R R L F W R P R L L Q K
CTCCCGGAGCTTGATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGG
GAGGGCCCTCGAACATATAGGTAAAGCCTAGACTAGTTCTCTGTCTACTCTTAGCAAAAGCTACTAATCTGTTCTACCTAACGTGCGTCCAAGAGGCC 600
A P G S L Y I H F R I . S R D R M R I V S H D . T R V I A R R F S G
CCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAAACAGACAATCGGCTGCTCTGATGCCGCGTGTTCGGCTGTCAGCGCAGGGCGCCCGGT
GGGAACCCACCTCTCCGATAAGCGGATACTGACCCGTGTGTCTGTTAGCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCGCGGGCCA 700
R L G G E A I R L . L G T T D N R L L . C R R V P A V S A G A P G
TCTTTTTGCAAGACCGACCTGTCCGGTGCCCTGAATGAATGACAGGACGAGGACGCGGCTATCGTGGCTGGCCACGACGGGCTCTCTGCGCAGCT
AGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCTCTGCTCCGTCCGCGGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGA 800
S F C Q D R P V R C P E . T A G R G S A A I V A G H D G R S L R S
GTGCTCGAGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTCCGGGGCAGGATCTCTGTCTATCTCACCTTGCTCTGCCGAGAAAG
CAGGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGATAACCCGCTTCACGGCCCCGTCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTC 900
C A R R C H . S G K G L A A I G R S A G A G S P V I S P C S C R E S
TATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTAC
ATAGGTAGTACCGACTACGTTACGCCCGGACGATGCGAATAGGCCGATGACGGGTAAAGCTGGTGGTTTCGCTTTGATAGCTAGCTCGCTCGTGCAATG 1000
I H H G . C N A A A Y A . S G Y L P I R P P S E T S H R A S T Y
TCGGATGGAAGCCGCTTGTGCTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTTCGCCAGGCTCAAGGCGCGCATGCC
AGCTTACCTTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTCGTAGTCCCGAGCGCGGTGGCTTGACAAGCGGTCCGAGTTCCGGCGCTACGGG 1100
S D G S R S C R S G . S G R R A S G A R A S R T V R Q A Q G A H A
GACGGCGAGGATCTCGTCGACCATGGCGATGCTGCTTGCCGAATATCATGTTGGAAAAATGGCCGCTTTTTCTGGATTTCGACTGTGGCCGGCTGG
CTGCCGCTCTTAGAGCAGCAGTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAGACCTAAGTAGCTGACACCGGCGGACC 1200
R R R G S R R D P W R C L L A E Y H G G K W P L F W I H R L W P A G
GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTAT
CACACCGCTTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCGCTTACCGACTGGCGAAGGAGCACGAAATSCCAT 1300
C G G P L S G H S V G Y P . Y C . R A V R R M G . P L P R A L R Y
CGCCGCTCCGATTTCGACGCGCATCGCTTCTATCGCTTCTTGACGAGTTCTTTCGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGAGGCCAA
GCGGCGAGGGCTAAGCGTCTCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAAGCTCGCCCTGAGACCCCAAGCTTACTGGCTGGTTCTGCTCGGGT 1400
R R S R F A A H R L L S P S . R V L L S G T L G F E M T D Q A T P

Tuesday, 18 November 1997 13:58
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 1

ACCTGCCATCACGAGATTTCGATTCCACCGCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGG
TGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGCGC 5100
N L P S R D F D S T A A F Y E R L G F G I V F R D A G W M I L Q R G
GGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCA
CCTAGAGTACGACC TCAAGAAGCGGGTGGGTTGAACAAATAACGTCGAATATTACCAATGTTTATTTTCGTTATCGTAGTGTTAAAGTGTTTATTTCGT 5200
D L M L E F F A H P N L F I A A Y N G Y K . S N S I T N F T N K A
TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
AAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACC GCATTAGTA 5300
F F S L H S S C G L S K L I N V S Y H V C I P S T S S . S L A . S
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGT
CCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTGTATGCTCGGCCCTTCGTATTTACATTTTCGGACCCACGGATTACTCA 5400
V S . L F P V . N C Y P L T I P H N I R A G S I K C K A V G A . . V
GAGCTAACTCACATTAATTGCGTTG
CTCGATTGAGTGTAATTAACGCAAC 5425
S . L T L I A L

Tuesday, 18 November 1997 13:56

Page 1

fig 53 pLM6 (1 > 4947) Site and Sequence

Enzymes : All 146 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

```
GTGGCAC TTTTCGGGAAATGTGCGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCGTGATAAT
CACCGTGAAAAGCCCTTTACACGCGCCTTGGGATAAACAATAAAAAAGATTATGTAAGTTTATACATAGGCGAGTACTCTGTATTGGGACTATTTA 100
V H F S G K C A R N P Y L F I F L N T F K Y V S A H E T I T L I N

GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTCCGTGTCGCCCTTATCCCTTTTTTGGCGCATTTCCTTCTGTTTTCCTGTCAC
CGAAGTTATTATAACTTTTCTCTCTCATACTCATAAGTTGTAAGGCACAGCGGGAAATAGGGAAAAAACGCCGTAAACGGAAGGACAAAAACGAGTG 200
A S I I L K K E E Y E Y S T F P C R P Y S L F C G I L P S C F C S

CCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
GGTCTTTGCGACCACTTTCAATTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTGCCATTCTAGGAACCTCAA 300
P R N A G E S K R C . R S V G C T S G L H R T G S Q Q R . D P . E F

TTCGCCCCGAAGAACGTTTCCAATGATGAGCACTTTTAAAGTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGCAAGAGCAACTCGGTCTG
AAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTTGAGCCAGC 400
S P R R T F S N D E H F . S S A M W R G I I P Y . R R A R A T R S

CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGACGTGCTGCC
GGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGCTTTTCGTAGAATGCCTACCGTACTGTCACTTCTTAATACGTACAGACGG 500
P H T L F S E . L G . V L T S H R K A S Y G V H D S K R I M Q C C

ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAA
TATTGGTACTCACTATTGTGACGCGCGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCTCTGATTGGCGAAAAACGTGTTGTACCCCTTAGTACATT 600
H N H E . . H C G Q L T S D N D R R T E G A N R F F A Q H G G S C H

CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
GAGCGGAAC TAGCAACCCCTTGGCCTCGACTTACTTCGGTATGGTTTGTGCTGCGCATGTGGTGCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA 700
S P . S L G T G A E . S H T K R R A . H H D A C S N G N N V A Q T

ATTAAC TGCGAAC TACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACCTCTGCGCTCGGCCCTTCCT
TAATTGACCGCTTGATGAATGAGATCGAAGGCGCGTTGTTAATTATCTGACCTACCTCCGCTATTTCACAGTCTTGGTGAAGACGCGAGCGGGGAAGGC 800
I N W R T T Y S S F P A T I N R L D G G G . S C R T T S A L G P S

GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCC TCCGTATCGTAC
CGACCGACCAATAACGACTATTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCACCATTCCGGAGGGCATAGCATC 900
G V L V Y C . . I W S R . A W V S R Y H C S T G A R W . A L P Y R S

TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACGTGACACCA
AATAGATGTGCTGCCCTCAGTCCGTTGATACCTACTTGTCTTATCTGCTAGCGACTCTATCCACGGAGTGACTAATCGTAACCATTGACAGTCTGGT 1000
Y L H D G E S G N Y G . T K . T D R . D R C L T D . A L V T V R P

AGTTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTTTGATAATCTCATGACCAAAATCCCT
TCAAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGTTTTAGGGA 1100
S L L I Y T L D . F K T S F L I . K D L G E D P F . . S H D Q N P

TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTTTCGCGCGTAATCTGCTCTTGCAAAA
ATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTTAGAAGAACCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTT 1200
L T . V F V P L S V R P R R K D O R I F L R S F F S A R N L L L A H
```

Tuesday, 18 November 1997 13:58
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 2

CAAAAAAACACCGCTACCAGCGGTGGTTTGGTTTGGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGCAGATACCAA
GTTTTTTTGGTGGCGATGGTGCACCAACAAACCGGCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTT 1300
K K T T A T S G G L F A G S R A T N S F S E G N V L Q Q S A D T I
TACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGA 1400
Y C P S S V A V V R P P L Q E L C S T A Y I P R S A N P V T S G C
GCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC
CGGTACCGCTATTACGACACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCAGTGTGTGCG 1500
C Q W R . V V S Y R V G L K T I V T G . G A A V G L N G G F V H T A
CCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGGGACAGGTATCC
GGTCGAACCTCGCTTGTGATGTGGCTTGACTCTATGGATGTGCGAC TCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTCCGCTGTCCATAGG 1600
O L G A N D L H R T E I P T A . A M R K R H A S R R E K G G Q V S
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTCTGACTT
CCATTGCGCGTCCCGAGCTTGTCTCTCGCGTGTCCCTCGAAGGTCCCTTTGCGGACCATAGAATATCAGGACAGCCAAAGCGGTGGAGACTGAA 1700
G K R Q G R N R R A H E G A S R G K R L V S L . S C R V S P P L T
GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAACACGCCAGCAACGCGGCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTG
CTCGCAGCTAAAAACACTACGAGCAGTCCCCCGGCTCGGATACCTTTTGGCGTCTTGGCGCGGAAAAATGCCAAGGACCGGAAACGACCGGAAAC 1800
A S I F V M L V R G A E P M E K R Q Q R G L F T V P G L L L A F C
CTCATGTCTCTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGATACCGCTCGCGCGAGCCGAACGACCGAGCG
GAGTGTAAGAAAGGACGCAATAGGGGAC TAAGACACCTATTGGCATAATGGCGGAAATCACTCGACTATGGCGAGCGGCTCGGCTTGTGGCTCGC 1900
S H V L S C V I P . F C G . P Y Y R L . V S . Y R S P Q P N D R A
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAACCGCTCTCCCCGCGCGTTGGCCGATTCAATATGCAGCTGGCAGCAGAGTTT
GTCGCTCAGTCACTCGCTCCTTCGCCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGTCGACCGTGTGTCCAA 2000
O R V S E R G S G R A P N T Q T A S P R A L A D S L M Q L A R Q V
CCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCGGCTCGTATGT
GGGCTGACCTTTGCGCCGCTACTCGCGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACA 2100
S R L E S G Q . A Q R N . C E L A H S L G T P G F T L Y A S G S Y V
TGTGTGAATTGTGAGCGGATAACAATTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAAGCT
ACACACCTTAACACTCGCTATTGTTAAAGTGTGCTTTGTGATACTGGTACTAATGCGGTTGCGCGGTTAATGGGAGTGATTCCCTTGTTTTCGA 2200
V V N C E R I T I S H R K Q L . P . L R Q A R N . P S L K G T K A
GGGTACCGGGCCCCCTCGAGGTGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCATGCAATGAGGAGGAGGAGCCAGAGA 2300
CCCATGCCCCGGGGGGAGCTCCAGCTGCCATAGCTATTGGAATATAGCTTAAGGACGTCGGGCCCTTAGGTACGTTTACTCCTCCTCCTCGGTCTCT
G Y R A P P R G R R Y R . A . Y R I P A A R G I H A N E E E E P E
U3 stuk

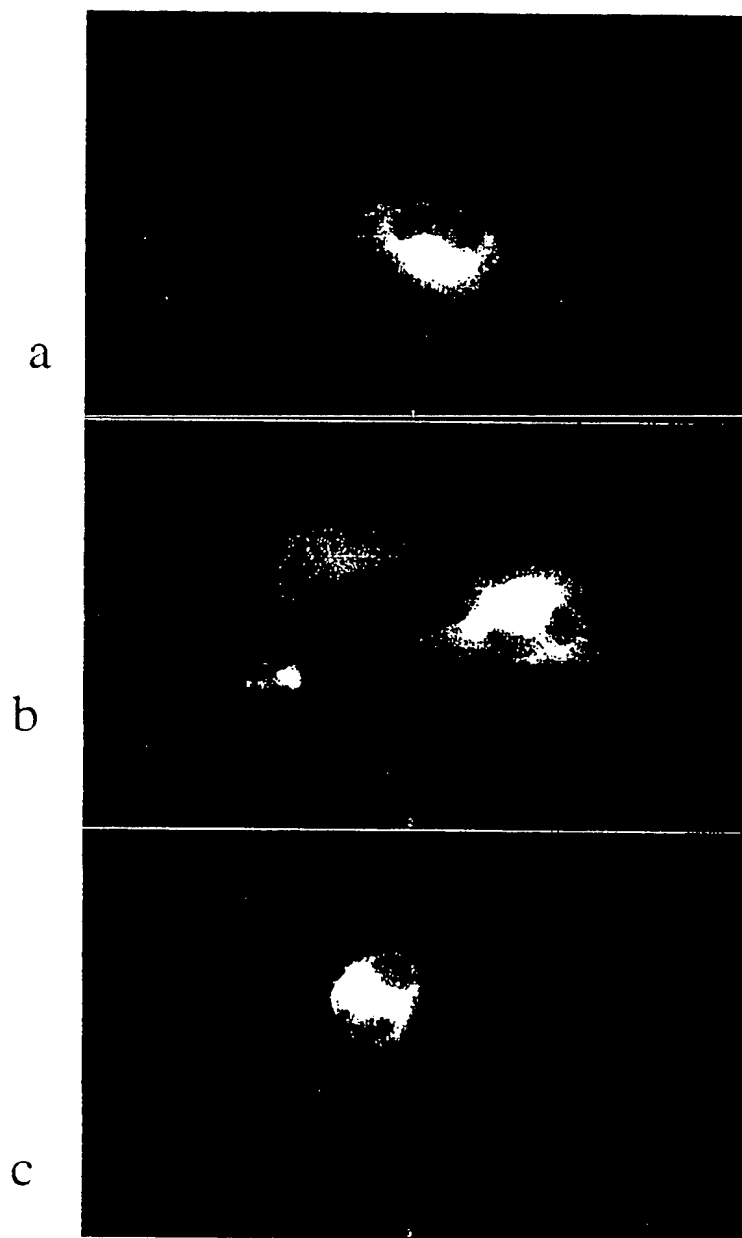


Figure 43

sonderdag, 27 november 1997 18:48

fig 44 pNP9 Map (1 > 10122) Site and Sequence

Enzymes : All 148 enzymes (No Filter)

Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

ATGACCATGATFACGCCAAGCTTGCATGCCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAGAAATTGGAGCAACTACC 100
CACATCCATTATGCCACCCGCGTTTCTAAGTGAGTTTAAATTTTGGTTTACGACTACAAAAATGTGTTCTTAAATACTATCTCGACTTGAGTCTATT 200
CTGTATGACTAGTTGTTGAGTGATTTCATTGAGAAAATATTAAGGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTAGTTTTTCGATC 300
AACTAGATCTTACAAAACCTTGAATACAATTCATTTTCAGATTACCTCGCCACGTGTCGCCACGTGAGCAACCGCTTCAGCAACTAACCCAAATTCGA 400
ACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCAGCTCAGAGTCAAGATATCGAAATTTGGTAAGAAATTTATTTGAGCTCAAACTTGATAAATGCC 500
CAGAAAGAGATGATAAAATGTAGTTTTTTTGCAAAACCTCCACCTTTATGCTCTAATATGACGGCTTATATCTCAATTTCTTGAGTTTTATCAAA 600
AAATTTTCCATATACAAATGTAGAAAAGTATTTTGCACAAATTTGTGAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTAA 700
AGTGGAAGGAGCGCAAGTCTATACTGGAATAATGATCTGAAACAAATTTGTGCTATTCTCAATGTTTAAGACATGTTTGAAGATTTTTCAAAATTCG 800
CACTAGTTTCAGAACCTTCTTTTGTATGAAAAAGTAAAAAAAACATTTTCAAACTCCACGCCACCATGTTTCAACTCTTAATTTTTATAAAATTT 900
TGCAATTTACAAATCGCTCCCTTGCCGAAAAGTGCCCAACCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATGATGTGAGAAAAACAGAAAG 1000
AGGCTAACTAAATGACAGGGACAGGTGTCT 1100
TTGTCCATTTTGTATTAACATTTGTGTGGAAGGAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTAATCTTTTCGGAAAT 1200
TGATGAATAAGATAGAGCCGATGACACTGGCTGGTAGTAGTATGAGTGAGAAATGCTTTTTCATCGTCTCAACTTGGCATGAGTCTTCCCCGCTCT 1300
CATCACTGACAAATTAATGTGCGGTTTTATGCGCTCTTCTCTATTCGGCACTCATTTCTGGGTACCAACAACTGGAATACATTTTACTACTATTCAGCC 1400
ATTTATTTTGATATTTAAATTTTGTCAATTAGGGATAAACACGCACTTTTAAAGTTTATTTAAAAAACGATATTTTCGATTTTAAAAATCTGAAAAGT 1500
TTCAAAAAATCAATAATATTCCTAACAAATTTGTATGGCTAAAAATTTTATTTCTACTGTTGACAAATCTTTATATGTATCACTGTTTCCATCTCAAA 1600
ACCTTGAATCCCCAAGTTATAGGAAGCTCCGTGTCAATTTCCATGCTATGAATCGCTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAA 1700
TTATTGGGCACGGTAATAAAGTCAAGCAGTTAGAATTTAATTCAAGCAGAGATTATCTCAATTCATCTTTGAACATTCAGCCAGTTCTGTACAA 1800
TTTTCCATGCTTTTGGCCATTAAAAAACTTTCTCACCTCTTCATCCATCTCACTCGTATCAAAAAATATAGCAAAAGCCCGACTCTACTTTTAAAG 1900
AGAAGGAGATACTGAGCCACATGGCGTGTGACCTTTTCT 2000
CTTGTTTCTTCTCTGTTTGACTCGGCTATTTTGTGCTGCTGAAAGCCGGGAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100
AAAAACGAGGCAATTTAAAAATATTAATAATGAGGTTGTAGATGTAGATTTGCAAAAGAGAAAAAAACAAAAAATAGGAACCGCCAGATCAAAA 2200
TTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTGGCTGAACAGAAACTGAAGTGCCTGCATAAATCTAGTGAACGTTTAGATTGAACCTGGAA 2300
ATCCTAAGCCTGAACATATAGCCTTATTCTAGATCTTAGTTGGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCAATTTAGTTTAAAGCTAGATTGACT 2400
GCTTGCTTCAGTCTAATCCAGACTAGATTTCCAGAGAGTTTTCAATTTTAAATGTTTCCAGTTTCTTGTACTTAAATCTTAATGCCCTGTGATGCGT 2500
AAAACTGTTATCCCTTTCTCTCACACTTTCAATTACAGATTCAATCAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAACACCCCTCATCATCAACC 2600
ACTTCATCAATAATACAAATTCATTCCGTCGTCGAGCGTTCGAGTGGAATAATAATGTTGGCTCGACGATATCCACATCTGCGGAAGAGCTTAGGTA 2700
TCCGATCCTTCCGGCTTCTTTTGAAGATATATTTTTCAGAAATCATCATCAACGTACAGCTCTATTTTGAATCTAAACCGACCTACCTCCCAACTCCA 2800
AAAACTCTAGACCAAAACCCAGCTAGTTGCTGTTGCTACAACTACAAAAATCGGAAGCTCAAGCTAGCTCGCTCTGAAAGCCGTGAGCACCCCAAAA 2900
CTTGCTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCGATAACAGCGGTGGTGGTGGTGGTGAATGCTGAAATTAAGTTATTCAGTAGCAAAACC 3000

dondorag, 27 november 1997 16:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence

Page 2

CATCTCTCATCGAATAGCCACAACCTACGAGAAAGCGCGGGTCCCTCAACAACAACTTTGTGAAAATCGCTGCCCAAGTGAAGTGGCT 3100
GAAGCGCCGACAGTAAGCTGGGAAGTGCACGTCTATGTGGAAGCTTTGTACGCCAAAAGTTCTACCGTAAAACGGACGCCCAATCATATCTCAA 3200
CAAGACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCAGCTGCCAACGTCATCATCGACGGAAGGTTCCCTAA 3300
GCATGCATTCCATCTTCCAAGAGTTCAAGTCAGACGAAAAGTCTCGTCATCAGACGATCTTACTCTTAACGCTCCATCGTGACAGTATCAGACA 3400
GCCGATAGCCGCAACACCGGTTCTCCAAATATTATCAACAAGCTGTTGAGGAAAAACCAACTGGCAGTGAAGGAGTGAAGGACACAGCGAAAAA 3500
GATCCACCTCAGCTGTTCCGCCAGTGACACCCAGCCAACATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAAGTTGACAAATGACCCCGTGATAT 3600
CTGAAAAACGAGAACCTGAAAAGTCCAATCAATGAGCATEGACACGAGGAGCTTCCACGCTTCCACTCTAAAATCAGTTGTTCCACTTAAAATGAC 3700
TTCAATCCGACAACCACCAACGTACGATGTTCTCTAAAACAAGGAAAAATCACATCGCTGTCAAGTCTTTGGATATGAGCAGTCTGCGGCTGTGAA 3800
GACTCCATTGTGGCTCATGCGTGGCTCAGTGACTCCGCGCAAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGATAAGACATCAG 3900
AATCCAGCGGCTACACTCTGACGCGGCTGTTGGGATGTGGCCAAAATCAGGAGAGCTGAAAGATACGATGACATGACTCGTGGAGCAGAAACGG 4000
CTATCTGACAACTTGAAGACAGTTCTCTTGTGCTGTGGAATATCCGATAACACAGAGCTCGACGACATATCCACGGAGGATTTGTCGGAGTAGAC 4100
ATGGCAACAGTGGCTCCAAACATAGCGACTATTCCTCACTTTGTTGCGCATCCACGCTTCTCTCTCAAGCCCCGAGTCCCAAGTGGTCTCCCAT 4200
CAGTCGATTTCTGATCTCGAGCAGAACAGGAGAATGTGTACAACTTCTGTCCAGTGGCGAACGAGCCAACGTGGCGCGCTGCCACCTCAACCTTCGG 4300
ACAACATTCGTAAGATCCCGGATCTCATCTATTCTCCACACTTATCAGTGTGACGTGATAAGGACACATGTCTATGCACTCACAGACTAGTGA 4400
CGAECTTCTTCAAAAACCAAGCTATTCAAGCCAATTTCACTCACTGATCGTAATGCCACCTTCAAGAGTTCACTCCACGAGCAGCAATGGCGG 4500
CTCTCTTGAGCCGAGAGGGTCCGAACTCGATGTGGAATATGATTCTTCAGGATCTACTCGCGCGTTCCGAGGTGGAAGCTCTACTGGTATCTA 4600
TGGAGAGAGCTTCAACTGCACAGACTATCCGATGAAAAATCCCCCGCATTCTGCCCCAAGTGAGATGGGATCCCACTATCACTGGCTAGCAGACA 4700
GCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTGGGACATGGCAGCTGACTGGAGTGTACAAGAACAATGTGCTACTAACAAGAAAC 4800
AGGAGAACTATGGAGCATTGTTGATCTTTTGAAGCAAGCTTAGAAAACCTCACTCAACATTCGATCGATCCAACTTGAAGCTGAAGAGGCAATACG 4900
ATTCAAGCAGGACATTTGCTATTTGAGGGATATTAGCAATCATCTTCATCCAACTCAGCTCATGTAAAGAGGCGCTGGTGAGCTTCTCTGTAACCA 5000
TCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCGTCGTGAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAACAGA 5100
AGAGCTGGATCCGCTCTCACTCTCCAAGTTCACCAAGAAGAACAACAAGAACTACGACGAAGCACAATATGCCATCAATTTCCGATCTCAAGGAACCT 5200
TGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAGAACGCGATAGTGCACTTTACGAAGTCCGCTTGACAATCTGGATCGTGCCCGCAAGTT 5300
GATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAGAAGAAGTGGACAACTCACCAAGGTCAGCCACTCGTGCTCTT 5400
CCCGCGCTCAATTCAGTTATCTACGACGATGAGCATGTCTATGATGACGCTGTAGCAGTACATCAGCTAGTCAATCTTGAAACGATCTCTGGCTG 5500
CAACTCAATCAAGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTGATCGTTAAACCCGGACAAAGAGATAATCTAGGATATCTTGCCATGTC 5600
ACCACTCAGTCATGCTGGAAGACATTGATGTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGACGACATCAACTTGAATCGATG 5700
CTCGTATCTATCTTGGCTATCAAATTTGGTGAACCTCGAEGCTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAACTGACATTTCTACTC 5800
CTCACTACAATCCGAATGTTGATGACGGTGGCGACAGAGTGGGTAGACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAATGATTTCTCAACTC 5900
GTCAAGTCAATTTTGACAGAGAGCTCTGGTGTAGCTGGAGCACTGGAATTGAAAGAGCAAACTGGCGAAGACCTGGCTGCTTATGTATCTATTC 6000

donderdag, 27 november 1997 16:48
fig 44 pNFP Map (1 > 10122) Site and Sequence

Page 3

GAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAACAATAAGAAGAAATTGCTTCAAGTGGAAACGACGCTCGAAAAGATCTTGAG 6100
AAGCAAGAATCATGCATCGTAATTCAGATAATATCCAAAGAATCGAATTGCATTTGTGTATCCGTTTTGCAAATGTCCCACTTCAAAAACACGAA 6200
GGTCCATTGTAGTATGCACAGTCAACCGATATCAATCCCTGAGCTTCAAAATCACCACAATTTCAAATGTCAGTAATGTGAATCGTCTCGAAGGAT 6300
TCATCTACGTTACCTCCGACGACGGGCGTAGAGGATGAGTATCGTCTAATGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAAT 6400
AGCTCTTCAGGCGCTCAATAATTTTATTGAGAAAACGAATTCGTGTGATGTGACAGTTGGTCCAAGAGCATGCTTGAACCTGCTCTAATGTCGATGGA 6500
TCCCGTGAATGGTTCAATCGATTGTGGAATGAGAACTTCATTCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAACCTTCGGTCTGCTGCACTTCT 6600
TCGAGGATCCCAACGACATCGTCTCTAAAAATGGCGGTGTTGATGGTGAACCCGAGAAATGTGCTCAAACTGCTTCACTTCAAGACCTCGTCCC 6700
GTCACCTGCCAACTCATCCGACAACACTTCAATCCCTCGAGTGTGATCCAATTGATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAA 6800
GACTCTAATCTTCTCTGCTCTCTCCCGCTTCTCTTATCTTGTACCGGTACCATGGTATTGATATCTGAGTCCCGCATCGGCGGTGTCATCAGATCG 6900
CCATCTCGCGCCGTGCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTCTCTCTCTGCTCCACCCCTATTTTTGTTATTAT 7000
CAAAAAAATCTTCTTCAATTTCTTTGTTTTAGCTTCTTTAAGTCACTCTAACAATGAATTTGTGTAGATCAAAAAAGAAATTAATTCGTAATA 7100
AAAGTCGAAAAAATTTGTGCTCCCTCCCCCATTAATAAATTCATCTCCAAATCTACACAATGTTCTGTGACACTTCTTATGTTTTTTTACTTCT 7200
GATAAATTTTTTTGAAACATCATAGAAAAACCGCACAAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCAG 7300
TCTGGGCTCTCATGACGTCAATCATGCTCATCGTGAAGGTTTTGGAGATTTTTGGAATTTTCAATCAAGTGAAGTTTATGAAATTAATTTTCC 7400
TGCTTTTGCTTTTGGGGCTTTCCCTATTGTTGTCAAGAGTTTCGAGGACGGCTTTTCTTGCTAAATCACAAGTATTGATGAGCACGATGAAGA 7500
AAGATCGGAAGAGGTTTGGGTTGAGGTCAGTGGAAGGTGAGTAGAACTTGATAATTTGAAAGTGGAGTAGTGTCTATGGGGTTTTGCTTAAATGA 7600
CAGAATACATTTCCCAATATACCAACATAACTGTTTCTAAGTACGCGGTACGGGCTTTCTGCTCGCGGCTTTCGGTGATGACGGTGAACCTCT 7700
GACACATGACGCTCCCGAGACGGTCACAGCTTGTCTGAAGCGGATGCCGGGAGCAGACAAGCCCTGACGGGCGGTGAGCGGTGTTGGCGGGTGTG 7800
GGGCTGGCTTAATATGCGGCATCAGAGCAGATTGTACTGAGAGTGACCATATGCGGTGTGAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATC 7900
AGGCGGCTTAAGGCGCTCGTGATACGCTATTTTTATAGGTTAATGTGATGATAAATGTTTCTTAGAGTCAAGTGGCACTTTTCCGGGAAATGTG 8000
CGCGGAACCCCTATTGTTTATTTTCTAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGA 8100
AGAGTATGAGTATTCACATTTCCGTGTCGCCCTATTCCCTTTTTGCGGCATTTTCCCTTCTGTTTTGCTCACCAGAAACGCTGGTGAAGTAAA 8200
AGATGCTGAAGATCAGTTGGGTGCAGAGTGGTTACATCGAAGTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTTTTGCCCCGAAGAACGTTTTCCA 8300
ATGATGAGCACTTTTAAAGTCTGCTATGTGGCGGGTATTATCCCGTATTGACGCGCGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAATG 8400
ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATTTACGGATGGCATGACAGTAAGAGAAATTATGAGTGTGCTAACCATGAGTGATAAAGTCTGC 8500
GGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTGCAACATGGGGATCATGTAATCTGCTTGTATGCTGGGAACCG 8600
GAGTGAATGAAGCATACCAACGACGAGGTGACACACAGTGTGATGAATGGCAACAGTTGCGCAACTATTAACTGGCGAACTACTTACTE 8700
TAGCTTCCCGCAACAATTAATAGACTGGATGGAGCGGATAAGTTGACAGCACTTCTGCGCTCGGCTTCCGGCTGGCTGGTTATTGCTGATAA 8800
ATCTGGAGCGGTGAGCGTGGGTCTCGGTATCATTGAGCACTGGGGCAGATGGAAGCCCTCCGATCTGATGTTATCTACAGCAGGGGAGTCAG 8900
GCACTATGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAGCATTGGTAAGTGTGACCAAGTTTACTCATATATACTTTAGA 9000

donderdag, 27 november 1997 16:48
fig 44 pNP9 Map (1 > 10122) Size and Sequence

Page 4

TTGATTTAAACTTCATTTTTAATTTAAAGGATCTASGTGAAGATCCTTTTGATAATCTCATGACCAAATCCCTTAACGTGAGTTTTCGTTCCACTG 9100
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTCTGCGCGTAATCTGCTGCTTGCAACAAAAAACACCCGCTACCAGCG 9200
GTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAATCTGCTCTCTAGTGAGCCGT 9300
AGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGECTACATACCTCGCTCTGCTAATCTGTTACAGTGCTGCTGECAGTGGCGATAAGTCGTCT 9400
TACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGCTCGGGCTGAACGGGGGTTCTGTGCACACAGCCAGCTTGAGCGAACGACCTAC 9500
ACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCAGCTTCCCGAAGGGAGAAAGGCGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG 9600
GAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCCTGGTATCTTTATAGTCTGTGCGGTTTCCGACCTCTGACTTGAGCGTCGATTTTTGTGATGCTC 9700
GTCAGGGGGGGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTCTCTCGCTTA 9800
TCCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAAGTATACCGCTCGCCGAGCCGAACGACCGAGCGCAGCGACTCAGTGAGCGAGGAAG 9900
CGGAAGAGCGCCCAATACGCAACCGCTCTCCCCCGCGGTTGGCCGATTTCATTAATGCAGCTGGCACGACAGGTTTCCGACTGCAAGCGCGGCACTGA 10000
GCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTATGCTTCCGGCTCGTATGTTGTGGAATTGTGAGCGGATAA 10100
CAATTTACACAGGAAACAGCT 10122

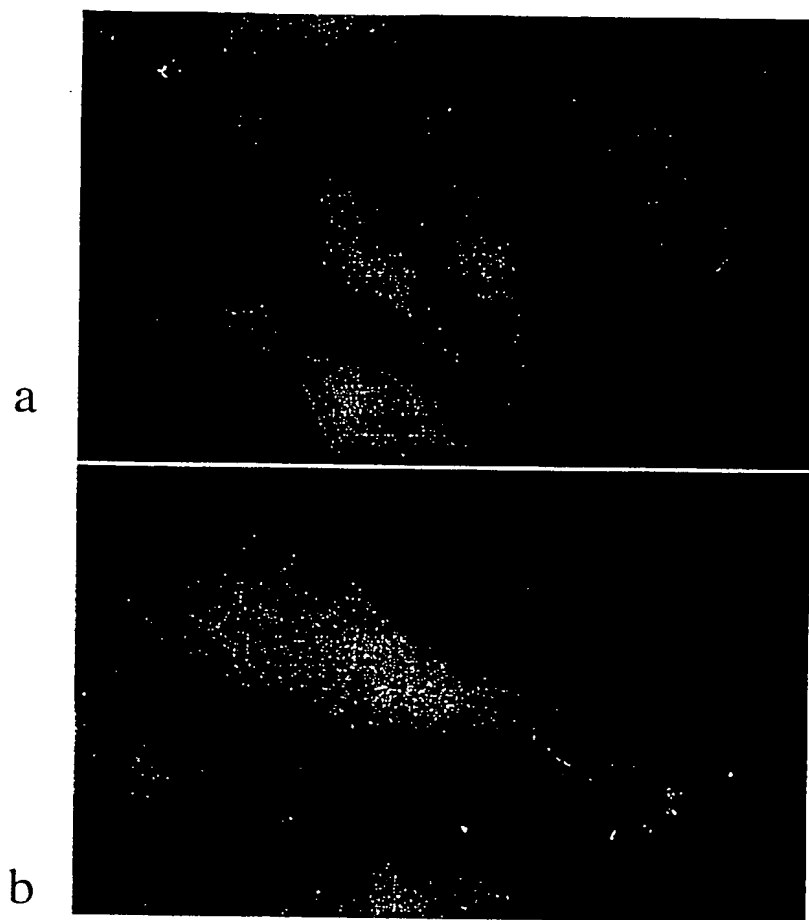


figure 45

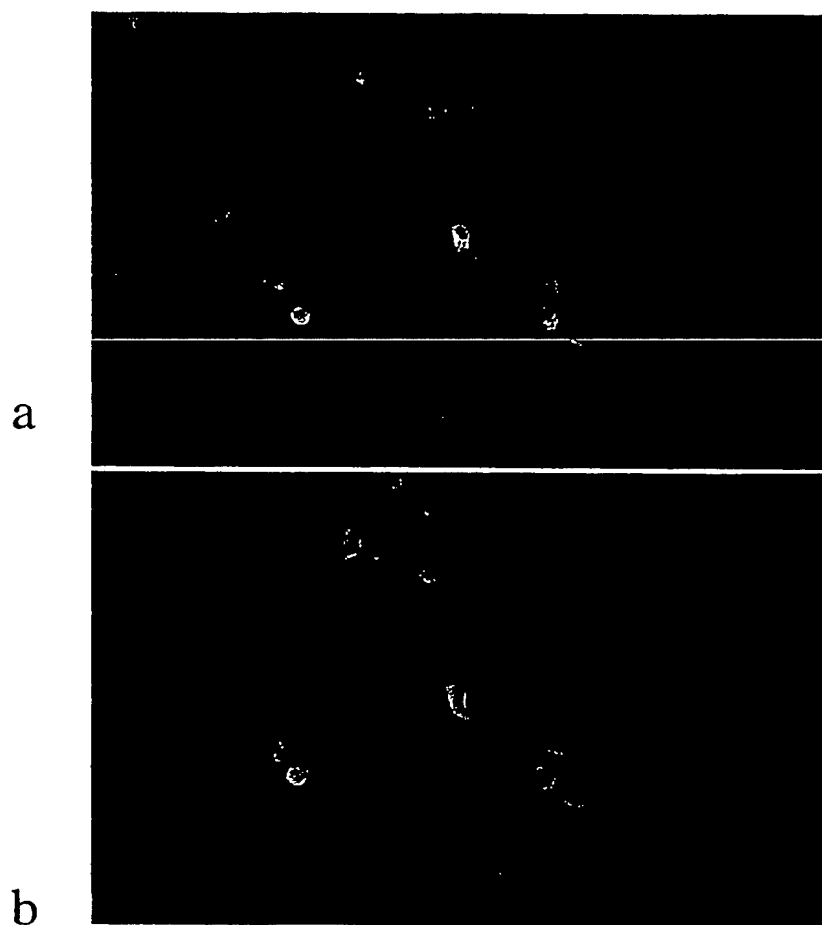


Figure 46


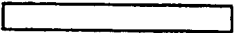

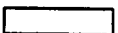





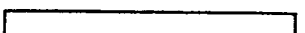

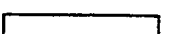

		+ end	MTB	actin
pTB72		+	+	+
pcDU2		nd	nd	+
pcDU3		nd	nd	+
pcDU4		nd	nd	+
peGFP72		+	+	nd
peGFPsma		- nd	+	
peGFPecf		-	+	nd
peGFPxba		-	-	nd
peGFPsac		-	-	nd
pLM4		+	+	nd
pCB251		nd	nd	nd
pLM5		nd	nd	nd
pCB201		nd	nd	+

Figure 47

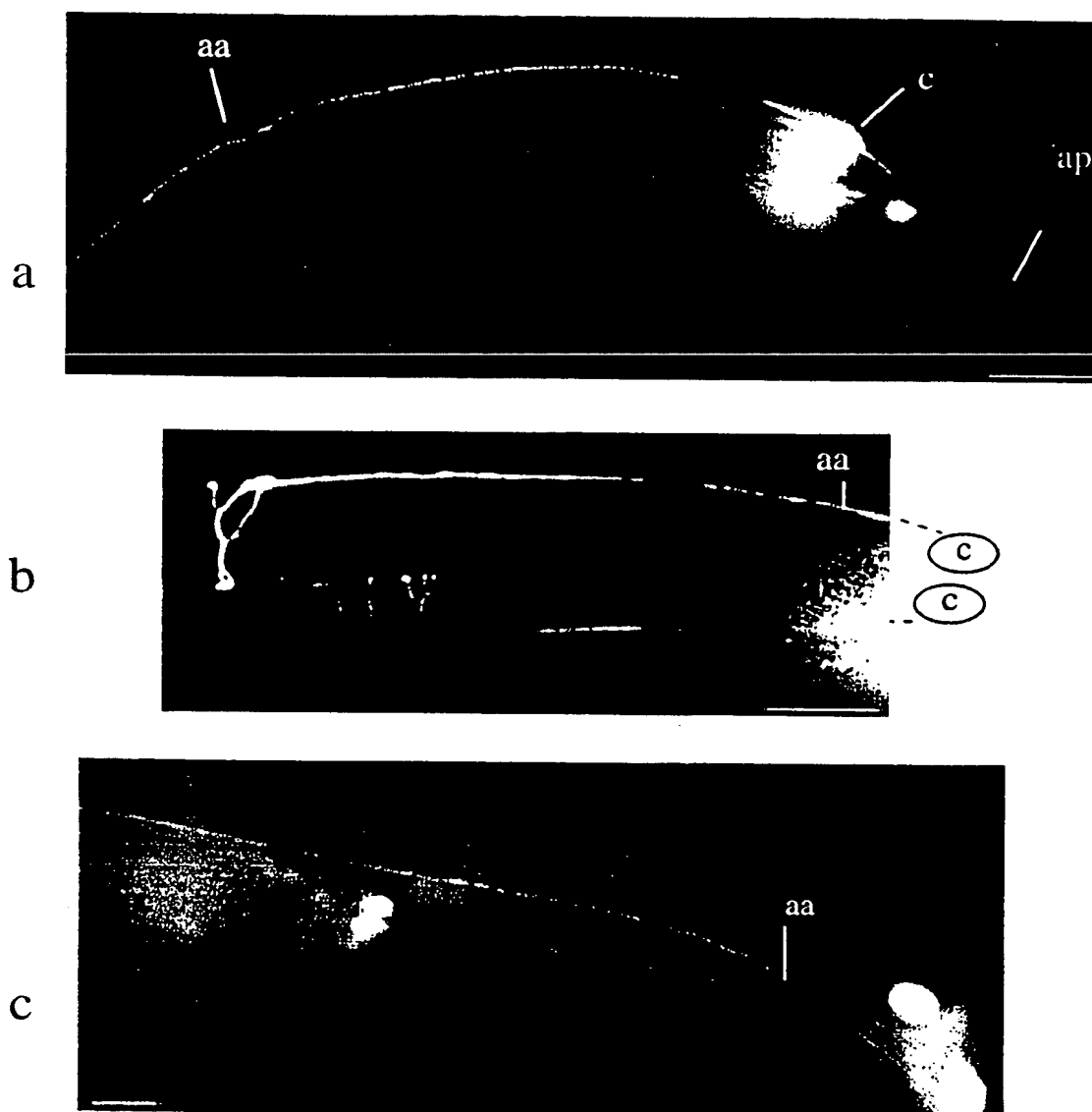


Figure 50

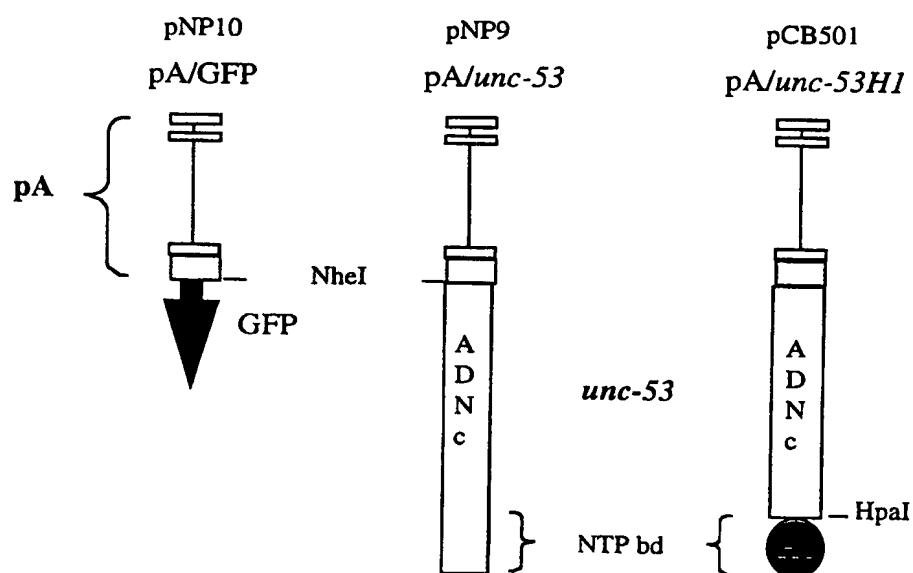


Figure 51a

donderdag, 27 november 1997 10:46
fig 25 pNP9 Map (1 > 12641) Site and Sequence

Page 3

GATTTTCAGACTTGGGCTATAAATTTTGTCAAACTAGGAATCTTAAATATTTGTATTTTTCGAAGAATGCTCTCAATCTCAAATTCATATTTTATA 6100
ATTTTCAGACCCCGTGATATCTGAAAAACGAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGAGCTTCACCGCTTCACCTCTAAATCAG 6200
TTGTTCACCTTAAATGACTTCAATCCGACACCAACCACTGACGATGTTCTTCTAAAAACAAGGAAAAATCAGATCGCTGTCAAGTCGTTTGCATATCA 6300
GCAGTCGTCGCGCTCTGAAGACTCCATTGTGCTCATGCGTCGGCTCAGGTGACTCCGCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATG 6400
GGAAAGAATAAGACATCAGAACTCAGCGGCTACACCTCTGACGCCGGTGTGCGATGTGCGCCAAAATGAGGGAGAAAGCTGAAAGAATACGATGACATGA 6500
CTGTCGAGCAGACAAACGGCTATCTGACAACTTCGAAGACAGTTCCTCTTGTCTCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGA 6600
CGATTTGTCCGGAGTAGACATGCAACAGTCGCTCCAAACATAGCGACTATTCCTCACTTTGTTCGCGATCCCACTCTTCTTCTCAAGGCCCGGAGTC 6700
CCCACTGGTCTCCACATCAGTCGATTCTGATCTCGAGCAGAACAGGAGAATGTGTACAACTTCTGTCCAGTGGCGAAGGAGCAACGTGGCGCCG 6800
CTGCCACCTCAACCTTCGGACAACTTCGCTAAGATCCCCGGGATACTCATCTATTCTCCACACTTATCAGTGYCAGCTGATAGGACACAATGTCTAT 6900
GCACTCAGAGACTAGTCGACGACTTCTTCAAAAAACCAAGCTATTCAGGCAATTTTCACTTCTGATCGTAAATGCCACTTCAAGAGTTCCATCTC 7000
ACCGAGCAGAGAATGGGGCTCTCTTGAAGCCGAGACGGGTGCCAACTCGATGTGGAATATGATTTCTCAGGATECTACTCGGCGCTTCCCGAGGTG 7100
GAAGCTCTACTGGTATCTATGGAGAGAGCTTCAACTGCACAGACTATCCGATGAAAAATCCCCGACATCTGCCAAAAGTGAGATGGGATCCCACT 7200
ATCACTGGCTAGCAGCAGCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCCGGACATGGCAGTGACTTGGAGTGTACAGAACACTGTG 7300
GACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTTGTTGATCTTTTGAAGCAAAAGCTTAGAAAACTCACTCAACACATTGATCGATCCAACTGA 7400
AGCCTGAAGAGGCAATACGATTCAGGAGGACATTTGCTCATTTGAGGGATATAGCAATCATCTTGATCCAACTCAGCTCATGCTAACGAAGGCGCTGG 7500
TGAGCTTCTTCGTCAACCATCTCTGGAATCAGTTGCATCCCATCGATCAGTGATGTCTGCTGCGTGAAAAGCAGCAAGCAGGAGAGATCAGTTGAGC 7600
TCGTTTGGCAAGAACAGAGAGCTGGATCCGCTCTCACTCTCCAAGTTCACCAAGAGAGAACAGAACTACGACGAGCAGATATGCCATCAATTT 7700
CCGGATCTCAAGAACTCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAGCAAGCGGATAGTGCACTTTACGAAGTCCGCTTGACAATCT 7800
GGATCGTCCCGCGAAGTTGATGTTCTGAGGAGAGCAGTGAACAGTTGAAACCGAGAACAGCAATTAAGGAAGAGTGGACAACTCAACCGGT 7900
CCAGCCACTCGTGCTTCTCCCGGCTCAATTCAGTTATCTACGACGATGAGCATGTCTATGATGCAAGCTGTAGCAGTACATCAGTACTCAATCTT 8000
CGAAACGATCTCTGGCTGCAACTCAATCAAGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTGATCGTTAACCCGGACAAAGAGATAATCGT 8100
AGGATATCTTGCATGTCAACAGTCAGTCATGCTGGAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAAATTGATGTGGAG 8200
CATCAACTTGGAACTGATGCTCGTATTCTATCTTGGCTATCAAAATGGTGAACTTCGACCGGTATTGGACATCCACAACCATGATAACCAAGCCATC 8300
CAACTGACATTCTACTCTCACTACAACTCCGAATGTTGATGACCGGTGCCGACAGAGTCGCTAGACAGTCTGGTCTTGATATGCTTCTTCCAAA 8400
GCAAAATGATTCTCCAACCTCGTCAAGTCAATTTTGACAGAGAGAGCTGCGTGTAGTGGAGCAACTGGAATTGGAAGAGCAAACTGGCGAAGACCCTG 8500
GCTGCTTATGTA?CTATTGAA?AAATCAATCCGAGATAGTATTGTTAATATCAGATTCTCTAAAACAATAAAGAGAAATGCTTCAAGTGGAAACGAC 8600
GCTGGAAAGACTCTTGAGAAGCAAGAAATCATGCATCGTAATTTAGATAATATCCAAAGAAATCGAATTGCAATTTGTTGTATCCGTTTTTGCAATGT 8700
CCCACTTCAAAACAAAGAGGTCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAATTCACCAAAATTTCAAAATGTCAAGTAATG 8800
TCGAATCGTCTCGAAGGATTCATCTACGTTACCTCCGACGACGGCGGTAGAGGATGAGTATCGTCTAAGTACAGATGCCATCAGAGCTTTCAAA 8900
TCATTGACTTCTTCCCAATAGCTCTTCAGGCGTCAATAATTTATTGAGAAACGAATTCGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAAGT 9000

donderdag, 27 november 1997 16:46
fig 35 pNP8 Map (1 > 12841) Site and Sequence

Page 4

TCCTCTAACTGTCGATGGATCCCGTGAATGGTTCATTGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTGCTAGAGATGGCAAAAAAAC 9180
TTCGGTCGCTGCACCTCTTCGAGGATCCACCGACATCGTCTCTAAAAATGGCCGGTTCGATGGTGAACCCGGAGAAATGTGCTCAAACTGCTTC 9200
AACTCAAGACCTCGTCCCGTCACTGCCAACTCATCCGACAACTTCAATCCCTCGAGTCTTGATCCAAATTCATGCTACCAAGCATCAGACCAT 9300
CGACAACATTTGAACAGAAGACTCTAATCTTCTCTGCTCTCCCCGCTTCTTATCTTCGTACCGGTACCATGGTATTGATATCTGAGCTCCGCATC 9400
GGCCGCTGTCATCAGATCGCATCTCGGCGCCGTGCTCTGACTTCTAAGTCAATTACTCTTCAACATCCCTACATGCTCTTCTCCCTGTGCTCCAC 9500
CCCCATTTTTGTTATTATCAAAAAAATCTTCTTAATTTCTTTGTTTTAGCTTCTTTAAGTCACTCTAACAATGAAATTGTGTAGATTCAAAAA 9600
TAGAATTAATTCGTAATAAAAGTCGAAAAAATGTGCTCCCTCCCCCATTAAATAATTCTATCCAAAAATCTACAAATGTTCTGTGTACACTTC 9700
TATGTTTTTTTACTCTGATAAATTTTTTTGAAACATCATAGAAAAACCGCACAAAAATACCTTATCATATGTTACGTTTTCAGTTTATGACCGCA 9800
ATTTTATTTCTTCGACGCTCTGGGCTCTCATGACGTCAAATCATGCTCATCGTAAAAAGTTTTGGAGTATTTTTGGAATTTTTCAATCAAGTGAAG 9900
TTTATGAAATTAATTTCTGCTTTTGTCTTTTGGGGTTTTCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCTTTTCTGCTAAAAATCACAAGTAT 10000
TGATGACGAGTGAAGAAGATCGGAAGAGGTTTGGGTTTGGGCTCAGTGAAGGTGAGTAGAAGTTGATAATTTGAAAGTGAGTAGTGTCTATG 10100
GGGTTTTGCTTAAATGACAGAATACATTCCTAATATACCAACATAACTGTTCTCTACTAGTGGCGCTACGGGCTTCTGCTCGCGGCTTTCGGT 10200
GATGACGCTGAAAACTCTGACACATGACGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCGGGAGCAGACAAGCCGTCAGGGCGGCTCAG 10300
CGGGTGTGGCGGCTGTGGGGCTGGCTTAATATGCGGCATCAGAGCAGATTGACTGAGAGTGACCATATGCGGTGTGAAATACCGCAGAGTGGCT 10400
AAGGAGAAAAATACCGCATCAGGGGCTTAAAGGCTCTGATACGCTATTTTTATAGTTAATGTCATGATAAATGTTTCTTAGAGCTCAGGTGG 10500
CACTTTTGGGGAATGTGCGGGAACCCCTATTTGTTATTTTTCTAAATACATTCATATATGATCCGCTCATGAGACAATAACCTGATAAATGCTT 10600
CAATAATTTGAAAAAGGAGATGAGTATTCAACATTTCCGTGCGCCTTATTCCTTTTTTGGGCATTTTGCCTTCTGTTTTGCTCACCAG 10700
AAACGCTGGTGAAGTAAAGATGCTGAAGATCAGTTGGTGCACGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAAAGATCTTGAGAGTTTTCG 10800
CCCCGAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGATTATTCCTGATTGACGCGGGCAGAGCAACTCGGTCGCGC 10900
ATACACTATTCTCAGAATGACTTGGTTGAGTACTACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGAGTGTGCCATA 11000
CCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACTGCGAGGACCGAAGGAGCTAACCGCTTTTTTGCAACAATGGGGATCATGTAACCTG 11100
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAAACGACGAGCGTGACACCAGATGCCTGTAGCAATGGCAACAGTTGCGCAAACTATTA 11200
ACTGGCGAACTACTTACTTAGCTTCCCGCAACAATTAATAGACTGGATGGAGCGGATAAAGTTGACAGGACCACTTCTCGCTCGGCCCTTCCGGCTG 11300
GCTGTTTTATTGCTGATAAATCTGGAGCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT 11400
CTACACGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAGCATTGGTAACGTGCAGACCAAGTT 11500
TACTCATATATCTTTAGATTGATTTAAACTTCATTTTTAATTTAAAGGATCTAGCTGAAGATCTTTTTGATAATCTCATGACCAAAATCCCTTAAC 11600
GTGAGTTTTGTTCCACTGAGCTCAGACCCGTAAGAAAGATCAAGGATCTTCTTGAGATCTTTTTTCTGCGGTAATCTGCTGCTTGCACAAACAA 11700
AAAACACCGCTACAGCGGTGGTTTTGTTTGGCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCCACATACCAATACT 11800
GTCCTTCTAGTGTAGCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAGCGCTACATACCTCGCTGTCTAATCTGTTACAGTGGCTGCTGCTCA 11900
GTGGCGATAAGTCTGTCTTACCGGTTGGAATCAAGACGATAGTTACCGGATAAGGCGAGCGGTGGGCTGAACGGGGGTTCTGTGCACACAGCCAG 12000

donderdag, 27 november 1997 18:48
fig 35 pNP8 Map (1 > 12641) Site and Sequence

Page 5

CTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGGCTCCCGAAGGAGAAAGGCGGACAGGTATCCGGTA 12100
AGCGGAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCAGGGGAAACGCTGGTATCTTTATAGTCCTGTGGGTTTCGCCACCTCTGACTTGAGC 12200
GTGGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGGCGCCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTGTCTCA 12300
CATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGCAGC 12400
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCAATACGCAAAACCGCTCTCCCGCGCGTTGGCGATTTCATTAATGCAGCTGGCAGCAGAGTTTCCCG 12500
ACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTCCGGCTCGTATGTTGTG 12600
TGGAATTGTGAGCGGATAACAATTTTCACAGGAAACAGCT 12641

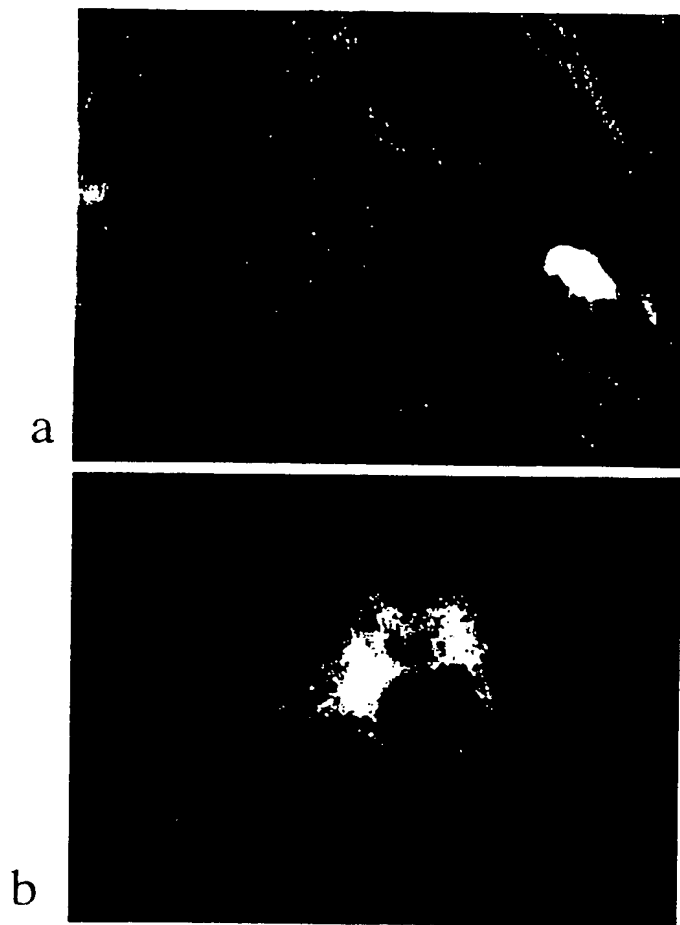


Figure 36: Association of *C. elegans* UNC-53 (expressed from pTB72) with the microtubular cytoskeleton of HepG2 cells. (A) microtubules stained with YL1/2 antibody to tubulin and (B) Staining for *C. elegans* UNC-53.

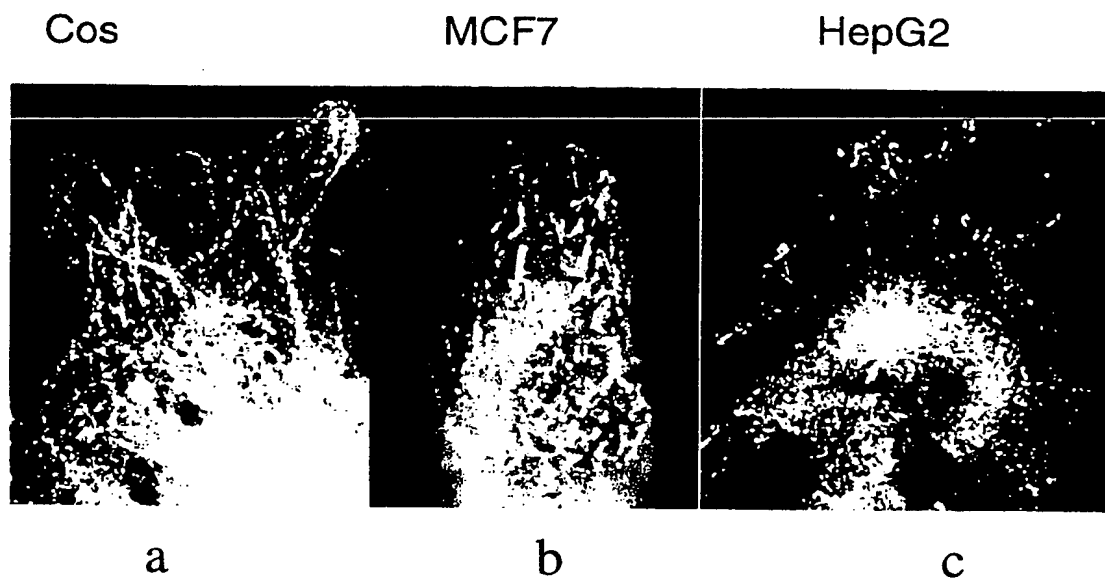


Figure 37: Microtubule (+)-end binding of *C. elegans* UNC-53 following transient transfection with pTB72 of HepG2 (a), MCF7 (b) and Cos cells (c). *C. elegans* UNC-53 was visualised by immunofluorescence using mab16-48.

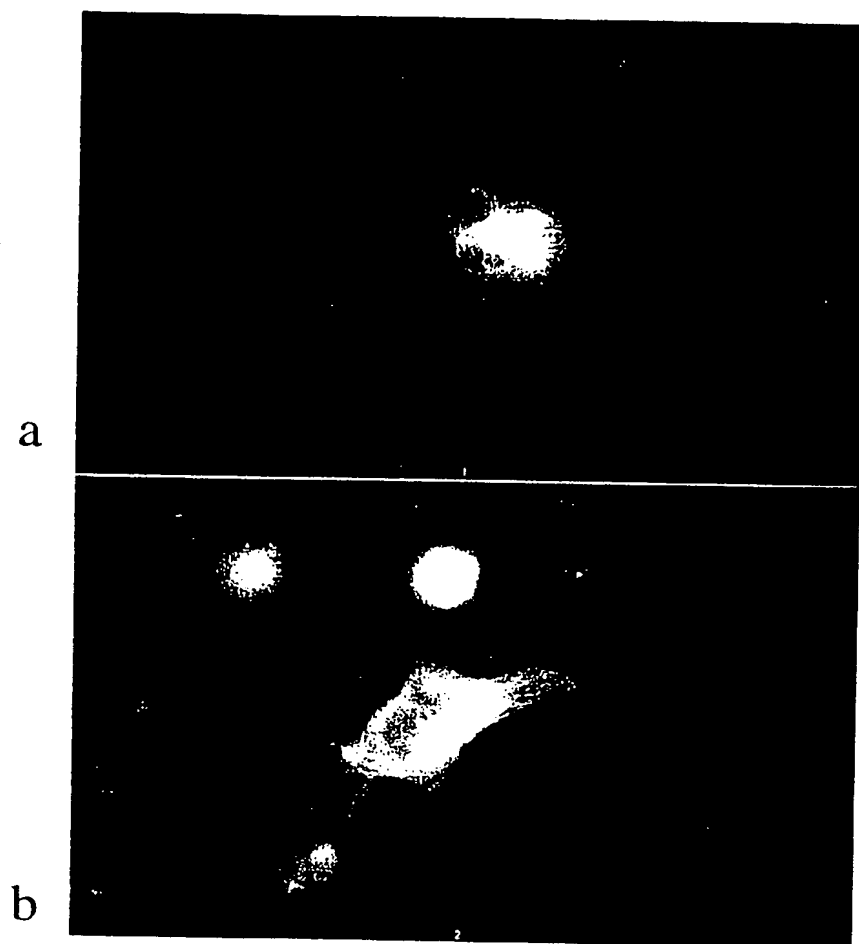


Figure 38

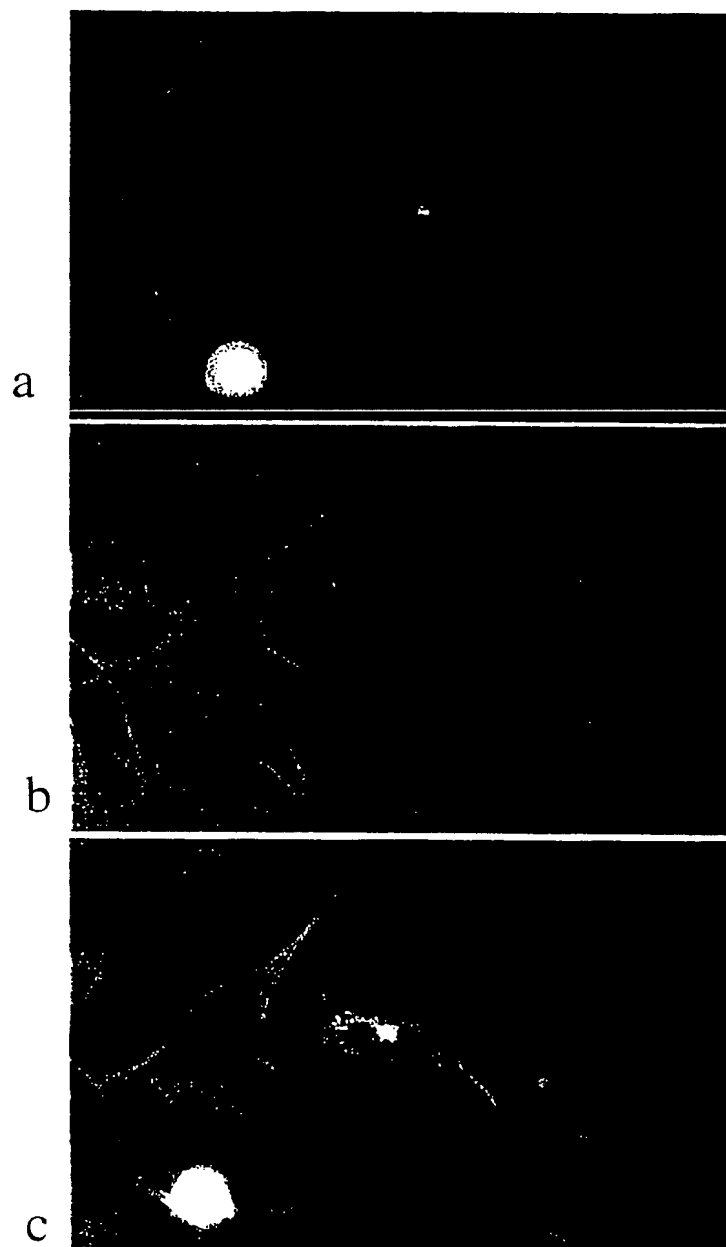


Figure 39

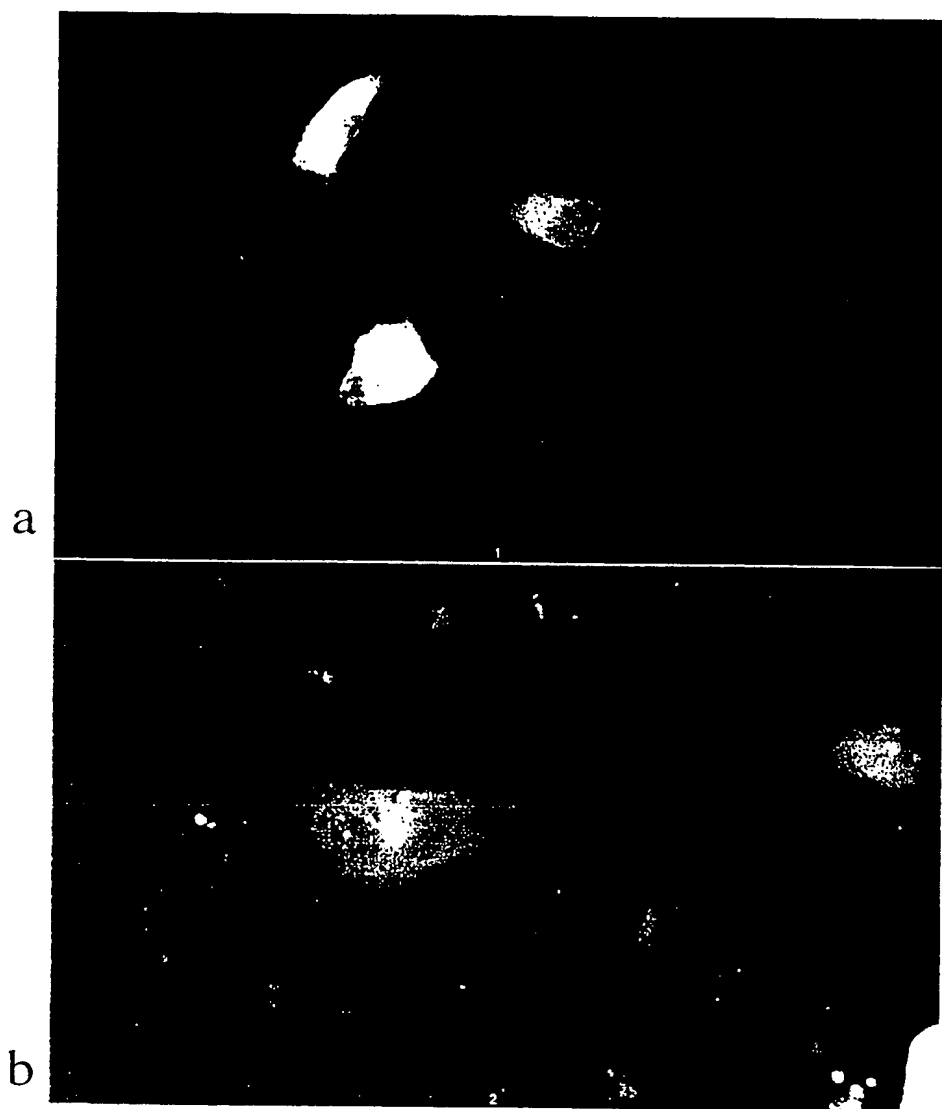


Figure 40

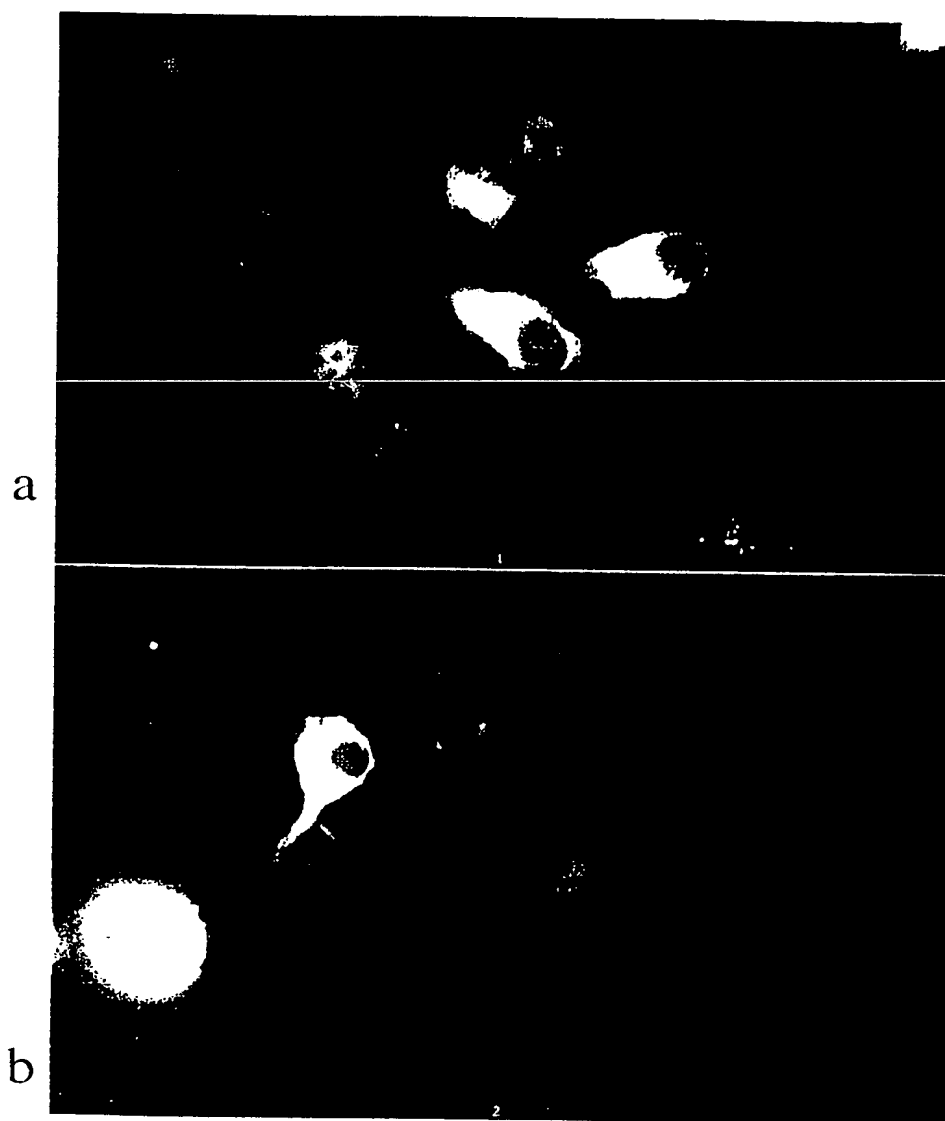


Figure 41

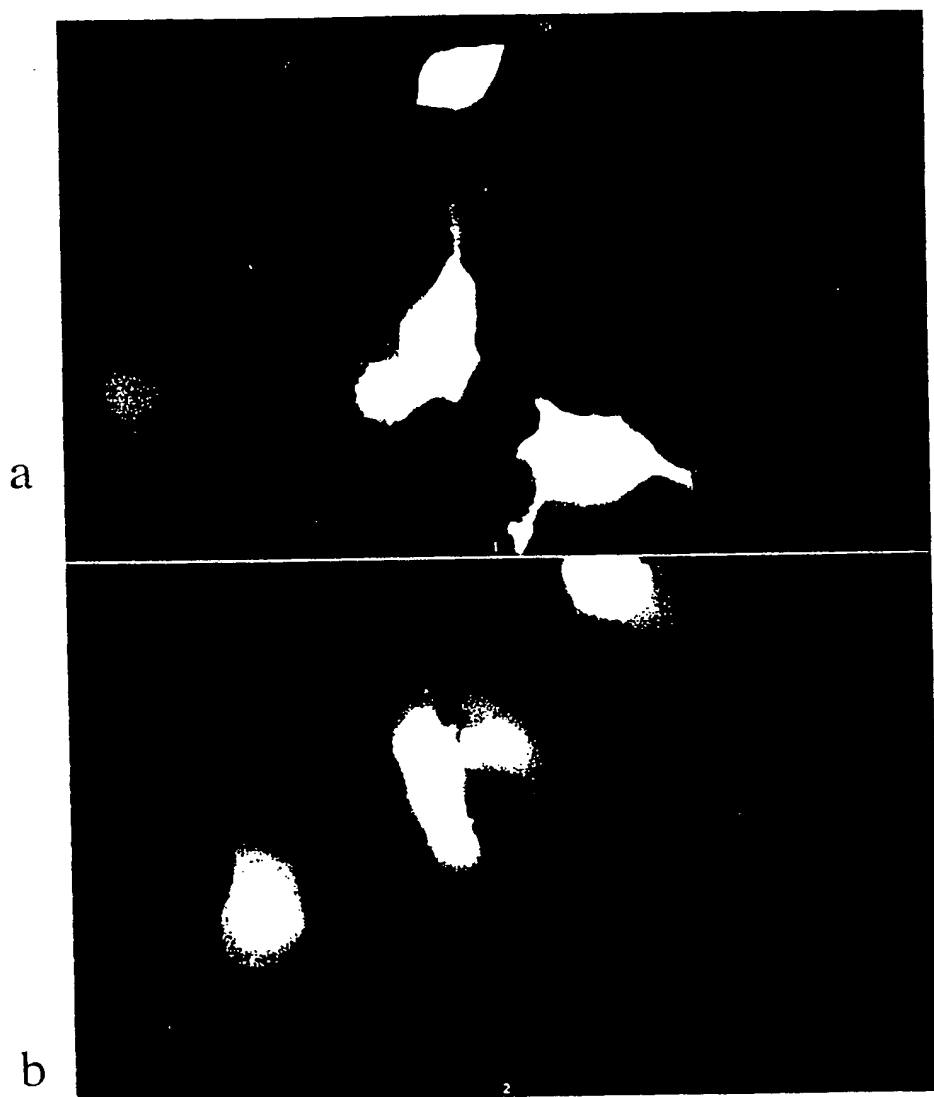


figure 42

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 7

CATAGACTTTCTGAAGAAAAGAACTCTGAGGCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACACACCCAAAGAACTTCGGATCAAGAGA
GTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTC 420

insert pLM1

ORF pLM1

I D F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R

CAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAAGA
GTTTGTAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTCTCTTTTCTTCT 430

insert pLM1

ORF pLM1

Q N S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K A

GTGGGTCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCGGATATAGAGGAGATTGCTAC
CAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGGTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATG 440

insert pLM1

ORF pLM1

S V V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T

ACCCGACTCTTCAGCCCCCTCATCCCCAACTACAGCATGGTTCCACAGAGACTGCTTACCCTCCATCAAGTCCCTCCACCTTGCTCTCGTGGGCACT
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGTCCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGA 450

insert pLM1

ORF pLM1

P D S S A P S S P K L O H G S T E T A S P S I K S S T L S S V G T

GATGTACCCGAGGGCCCTGCTCACCCAGCCCCCACTAGGCTGTTCATGCAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGSAGCTGCGGT
CTACASTGGCTCCCGGGACGAGTGGGTGCGGGGGTGATCCGACAAGGTACGTTTACTCTCTCTCGGTCTCTTCTCTCCATAGCCTCGACGCGA 460

insert pLM1

ORF pLM1

D V T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R

CTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACAT
GACTCGATACCCCTTCTCTTTACTTCAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCAAGCCCTCTGGTACGTGTGTGA 470

insert pLM1

ORF pLM1

S E L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T I H N F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

GCAGTTGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCT
CGTCAACCTCCACC TGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCAGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGA

insert pLM1

ORF pLM1

D L E V D L L K A E N D R L K V A P G P S S G S T P G Q V > G S 3

GCATTATCTTCCCCACGCCCTCCCTAGGCCTGGCACTACCCATTCTTTCGGCCCCAGTCTTGCAGACACAGACCTGTACCCATGGATGGCATCAGTA
CGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCGGGGTCAGAACGTCGTGTCTGGACAGTGGGTACCTACCGTAGTCAT

insert pLM1

ORF pLM1

A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S

CTGTGGTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATCTTCTTGGG
GAACACCAAGGTTTCTCTTCACTGGGAGGCCACCACTCTTACGGGGCGTCGTGTAGTAGTTTCCCTGAACCTCGTCGTCTTAAAGAGGACCC

insert pLM1

ORF pLM1

T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G

CTGTAGCAAGGTCAAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAGGACTATATTTCTAAATGGACCCAGCCTCTACC
GACATCGTTCCAGTCACCTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGG

insert pLM1

ORF pLM1

C S K V S G K V D W K M L D E A V F Q V F K D Y I S K M D P A S T

CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAACAGAGTGTGGATGCAGAGCCCCCGAGATGCCTCCTTGGCGTCGAGGTG
GACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCAC

insert pLM1

ORF pLM1

L G L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G

TCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACACCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACAT
AGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTACGCAGCTGTGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTA

insert pLM1

ORF pLM1

V N N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 9

```

AASCC TCCTGCTGAAGCACCGCGCCTCGTCC TCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACC TGGTGGAGCGC
TTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTTCGCCGTGCCCGTTC TGGATGGACTGGTTAGCGAACC GSC TCATGACCACCTCGCG
540
-----insert pLM1-----
-----ORF pLM1-----
S L L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R
TCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAAG
AGACCGGCACCTCAGTGCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCC TAGACGTGACATAGAAAGGTTGGATCGGTTGGTC
560
-----insert pLM1-----
-----ORF pLM1-----
S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q
TAGACCGGAAACAGGAATTGGGGATGTGCCCTGGTGATTCATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC
ATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTG
580
-----insert pLM1-----
-----ORF pLM1-----
I D R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T
CTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC
GACGTTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTAC TGTGGGTTGGTACCGAACGTGAAC TCGAAGTCCTACAAC TGG
570
-----insert pLM1-----
-----ORF pLM1-----
C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T
TCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACGAAGAGGC
AASAGGTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTTCTCC
580
-----insert pLM1-----
-----ORF pLM1-----
F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E
TGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCAGACACCTCAGACTTCCTCATCGGCCCTTGCTTCTT
ACGAAGCCACGAGCTGACCATGGGTTTCGACACCATAGTAGAGGTGTGGAAGGAAC TCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAA
590
-----insert pLM1-----
-----ORF pLM1-----
L L R V L D V V P K L V Y H L H T F L E K H S T S D F L I G P C F F

```

Page 10

[illegible]

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 11

ACAAACTCCTGGGCTTTCTGGGGAGGGGTTGAGAAACATCAAAACACTGACAGCAGTTCCCCGGAATTCAGCTTGGACTTAACCAGGCTGAACCTGCTCA
TGTGTTGAGGACCCGAAAGACCCCTCCCAAGTCTTTGTAGTTTGTGACGTCGTCAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGT 570

insert pLM1

T N S W A F W G G V Q K T S K H C S S S P E F S L D L T R L N L L

AAAGAAGCCGAATTCAGCACACTGGCGGCCGTTACTAGTTCTAGATAAATGATCATAATCAGCCATACCACATTTGTAGAGGTTTACTTGTCTTAAAA 580

TTTCTTCGGCTTAAGGTCGTGTGACCGCCGCAATGATCAAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTT

insert pLM1

K R S R I P A H W R P L L V L D N . S . S A I P H L . R F Y L L . I

AACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAAATAAGCAATA 590

TGGAGGGGTGGAGGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCTGTTAT

T S H T S P . T . N I K . M Q L L L L T C L L Q L I M V T N K A I

GCATCACAATTTACAAATAAAGCATTTTTTCTACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTAACGCGTAAATTGAAGCGTTA 700

CGTAGTGTTAAAGTGTTTATTTCTGTAAGGAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTCCGAAT

II ori

A S Q I S Q I K H F F H C I L V V V C P N S S M Y L N A . I V S V

ATATTTTGTAAAAATTCGGCTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGATAGAC 710

TATAAACAATTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCCGTTTAGGGAATATTAGTTTCTTATCTG

II ori

N I L L K F A L N F C . I S S F F N Q . A E I G K I P Y K S K E . T

CGAGATAGGGTTGAGTGTTGTCCAGTTTGGAAACAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACGCTCTATCAGGGGAT 720

GCTCTATCCCAACTCACAACAAGGTCAACCTTGTCTCAGGTGATAATTTCTGCACCTGAGGTTGCAGTTTCCCGCTTTTGGCAGATAGTCCCGCTA

II ori

E I G L S V V P V W N K S P L L K N V D S N V K G R K T V Y Q G D

GGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGAGCCCCGATTTAGAGCTT 730

CCGGGTGATGCACCTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGTGATTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAA

II ori

G P L R E P S P . S S F L G S R C R K A L N R N P K G S P R F R A

GACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGCAAGCAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGATAGCGGTACGCTGCGCGTAAC 740

CTGCCCCCTTCGGCCGCTTGACCGCTCTTTCCCTCCCTCTTTCTGCTTTCTCGCCCGCATCCCGCGACCGTTACATCGCCAGTGCGACGCGCATTG

II ori

R G K P A N V A R K E G K K A K G A G A R A L A S V A V T L R V T

CACCACACCCGCGCGCTTAATGCGCCGTACAGGGCGGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATA 750

GTGGTGTGGGCGGCGGAATTACGGCGGATGTCCCGCGCAGTCCACCGTGAAAGCCCTTTACACGCGCTTGGGGATAAACAATAAAAGATTAT

II ori

T T P A A L N A P L Q G A S G G T F R G N V R G T P I C L F F . I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 12

CATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCAGAAAGAACAGCTGTGGAA
GTAAGTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTATTATAACTTTTCCTTCTCAGGACTCCGCCCTTCTTGGTCGACACCTT
H S N M Y P L M R Q . P . . M L Q . Y . K R K S P E A E R T S C G 760

TGTGTGTCAGTTAGGGTGTGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGAAAGTCC
ACACACAGTCAATCCCACACCTTTCAGGGGTCGGAGGGTCGTCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGG
M C V S . G V E S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P 770

CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGC
GGTCCGAGGGGTCGTCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGGATTGAGGCG
Q A P Q Q A E V C K A C I S I S Q Q P . S R P . L R P S R P . L R 780

CCAGTTCGCCCATTTCTCGCCCATGGCTGACTAATTTTTTTTATTTATGAGAGGCGAGGCGCCTCGGCTCTGAGCTATTCCAGAAGTAGTGAGG
GGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAATAAATAAATACGTCTCCGGCTCCGGCGGAGCGGAGACTCGATAAGGTCTTCATCACTCC
P V P P I L R P M A D . F F L F M Q R P R P P R P L S Y S R S S E 790

AGGCTTTTTGGAGGCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
TCCGAAAAAACCTCCGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTAGCAAAGCGTACTAATCTGTCTACCTAACGTGCGTCCAAGAGG
E A F L E A . A F A K I D O E T G . G S F R M I E Q D G L H A G S P 800

GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCGTGTTCGGCTGTCAGCGCAGGGGCGCCG
CCGGCGAACCCACCTCTCCGATAAGCCGATCTGACCCGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCCGCTCCCGCGGGC
A A W V E R L F G Y D W A Q Q T I G C S D A A V F R L S A O G R P 810

GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAAGTGAAGACGAGGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTTGCGCAG
CAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCTC
V L F V K T D L S G A L N E L Q D E A A R L S W L A T T G V P C A 820

CTGTGCTCGACGTTGTCTGTAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCTGTCTATCTACCTTGCTCTGCGGAGAA
GACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCGCTTCACGGCCCCGCTCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTT
A V L D V V T E A G R D W L L L G E V P G Q D L L S S H L A P A E I 830

AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGT
TCATAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAACTAGGCCGATGGACGGTAAGCTGGTGGTTGCTTTGTAGCGTAGCTCGCTCGTGCA
V S I M A D A M R R L H T L D P A T C P F D H Q A K H R I E R A R 840

ACTCGGATGGAAGCGGCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAC TGTTCGCCAGGC TCAAGGCGAGCATGC
TGAGCTACCTTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTGCTAGTCCCCGAGCGCGGTGGCTTGACAAGCGGTCGAGTTCCGCTCGTACG
T R M E A G L V D Q D D L D E E H Q G L A P A E L F A R L K A S M 850

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 13

CCGACGGCGAGGATCTCGTCGTGACCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGC- 860
GGCTGCCGCTCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGCGAAAAGACCTAAGTAGCTGACACGGCCGA
-----Kan/Neo-----
P D G E D L V V T H G D A C L P N I M V E N G R F S G F I O C G R L
GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGG- 870
CCCACACCGCTTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCAGCGCTTACCCGACTGGCGAAGGAGCAGAAATGCCA
-----Kan/Neo-----
G V A D R Y Q D I A L A T R D I A E E L G G E W A D R F L V L Y G
ATCGCCGCTCCCGATTTCGACGCGCATCGCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC- 880
TAGCGGGCAGGGCTAAGCGTCGCTAGCGGAAGATAGCGGAAGAAGTCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTGGCG
-----Kan/Neo-----
I A A P D S Q R I A F Y R L L D E F F . A G L V G S K . P T K R R
CAACCTGCCATCAGGATTTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGACGCCGGCTGGATGATCTCCAGCGC- 890
GTTGGACGGTAGTGCTCTAAGCTAAGGTGGCGGGCGGAAGATATTTCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGG
P T C H H E I S I P P P P S M K G V A S E S F S G T P A G . S S S A
GGGGATCTCATGCTGGAGTTCTTCGCCACCCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA- 900
CCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTATGGCTTCTTGGGCGCGATCTGCCGTTATT
G I S C V S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .
AAAGACAGAATAAACCGCACGGTGTGGGTGCTTTGTTTCATAAACCGGGGTTCGGTCCCAGGGCTGGCACCTGTGCGATACCCACCGAGACCCCATTC- 910
TTTCTGTCTTATTTTGGTGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCAGCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC
K D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I
GGGCCAATACGCCCCGCTTCTTCTTTTCCCCACCCACCCCAAGTTCCGGTGAAGGCCAGGGCTCGCAGCCAACGTCGGGGCGGCGAGCCCTGCC- 920
CCCGGTTATGCGGGCGCAAGAAGGAAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCAGCGTCGGTTGACCCCCGCCGTCCGGGACGG
G A N T P A F L P F P P P T P Q V R V K A Q G S Q P T S G R Q A L F
ATAGCCTCAGGTACTCATATATATTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCA- 930
TATCGGAGTCCAATGAGTATATGAAATCTAACTAAATTTGAAGTAAATTTAAATTTTCTAGATCCACTTCTAGGAAAACTATTAGAGTACTGGT
P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . F
AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTG- 940
TTTAGGGAATTGCACTCAAAGCAAGGTGACTCAGTCGAGTCGGGGCATCTTTCTAGTTTCTAGAGAAGCTCTAGGAAAAAGACGCGCATTAGACGAC
-----pUC ori-----
K S L N V S F R S T E R Q T P . K R S K D L L E I L F F C A . S A
CTTGCAACAAAAAACCCCGCTACCAGCGGTGGTTGTTTGGCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCA- 950
GAACGTTTGTTTTTTGGTGGCGATGGTGGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGCTCGCGTC
-----pUC ori-----
A C K Q K N H R Y Q R V F V C R I K S Y Q L F F R R . L A S A E R F

• Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 14

ATACCAAATACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCTGTTACCA3
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTC TTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTC 960
pUC ori
Y Q I L S F . C S R S . A T T S R T L . H R L H T S L C . S C Y Q
TGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGTGAACGGGGGTTTCGTG 970
ACCGACGACGGTCACCGCTATTACGACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCATTCCGCGTCCGACGCCGAC TTGCCCCCAAGCAC
pUC ori
V L L P V A I S R V L P G W T Q D D S Y R I R R S G R A E R G V R
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGAC 980
GTGTGTCGGGTCGAACCTCGCTTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCGCGCTG
pUC ori
A H S P A V S E R P T P N . D T Y S V S Y E K A P R F P K G E R R T
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC 990
TCCATAGGCCATTTCGCCGTCCAGCCTTGCTCTCGCGTGC TCCTCGAAGGTCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG
pUC ori
G I R . A A G S E Q E S A R G S F Q G E T P G I F I V L S G F A T
TCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTACGGTTCTTGCCCTTTTGCTG 1000
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGTTCGTTGCGCGGAAAAATGCCAAGGACCGGAAAAACGAC
pUC ori
S D L S V D F C D A R Q G G G A Y G K T P A T R P F Y G S V P F A
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 10070
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
G L L L T C S F L R Y P L I L W I T V L P P C I

onderdag, 27 november 1997 16:46

fig 35 pNPB Map (1 > 12541) Site and Sequence

Enzymes : All 146 enzymes (No Filer)

Settings : Circular, Certain Sites Only, Standard Genetic Code

ATGACCATGATTACGCCAAGCTTGCATGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAATTGGAGCAACTACC 100

CACATCCATTATGCCACCCGCGTTTCTAAGTGAGTTAATTTTGTGTTTACGACTACAAAAATGTGTTCTTTAATACTATCTTCGACTTGAGTCTATT 200

CTGTATGACTAGTTGTTGAGTGATTTTTCATTGAGAAAAATTAAAGGAACATTATTTACTTTGCTTATTTGCCCTAAGTTGATTTAGTTTTTCGATC 300

AACTAGATCTTACAAAACCTTGAATACAATTCATTTTCAGATTACCTCGCCACGCTGTCGCCACGTGAGCAACCGCTTCAGCAACTAACCCAAATTCGA 400

ACTTTCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAAATATCGAAAATTGGTAAGAATTTTATTTTGAGCTCAAACTTGATATAAATGCC 500

CAGAAAAGAAGATGATAAAATGTAGTTTTTTTGCAAAACCTCCACCTTTATGCTCTAATATGACGGCTTATATCTCAATTTCTTGAGTTTTATCAAA 600

AAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTGTCAAGTTGACAGCTTTGTAATAGATCCAAATGGAACTAGATACAAGCTGTAA 700

AGTGGAGGAGCGCAAGTCTATCTGGAAATATGATCTGAAACAAATTTGTCTATTCTCAATGTTTAAGACATGTTTTGAAGATTTTTCAAAATTCG 800

CACTAGTTTCAGAACCTTCTTTTGTATGAAAAAGTAAAAAAAACATTTTCAACCTCACEGCGCCACCATGTTTCAACTCTTAATTTTATAAAATTT 900

TGCAATTTTCAAAATCGCTCCCTTTGCGCGAAAAGTGCCCAACAAATCAATTTCTCGGCTTCATAATGACTTTTAAATGATGTGAGAAAACACAGAAG 1000

AGGCTAACTAAATGACAGGGACAGGTGTCCCTCTTCTCCCTCTTCTCCCGCTCTCTCCCGTTCATCTCCAACCAACAAATTTTCAATTTTCG 1100

TTGTCCATTTTCTTATAAACATTTGTGTGTGGAAGGAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTAATCTTTTGGAAAT 1200

TGATGAATAAGATAGAGCGGATGACACTGGCTGGTAGTAGTATGAGTGAGAAATGCTTTTTCATCGTCTCAACTTGGCGATGAGTCTTCCCGCTCT 1300

CATCACTGACAATTAATGTGGGTTTTATGCGCTCTTCTCTATTCGCGCACTCATTTCTGGGTACCACAACTGGAATACATTTTACTACTATTCAAGCC 1400

ATTTATTTTGATATTTAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAGTTTATTTAAAAAACGATATTTTTCGATTTTAAAAATCTGAAAAGT 1500

TTCAAAAAATCAATAATATTECCTAACAAATTTGTATGGCTAAAAATTTTATTTCTACTGTTGACAAATCTTTATATGTATCACTGTTTCCATCTCAA 1600

ACCTTGAATCCCCCAAGTTATAGGAAGCTCCGTGTACATTTCCATGCTATGAATCGTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAA 1700

TTATTTGGGACCGGTAATAAGTGCAAGCAGTTAGAAATTTAATTCAGCACAGATTATCTATCAAATCAATCTTTGAACATTGAGCAGTTCTGTACAA 1800

TTTTCCATGCTTTTGGCCATTAAAAAATTTCTCACCTCTTCATCATCTCACTCGTATCAAAAAAGTATAGCAAAAGCCCGACTCTACTTTTTAAG 1900

AGAAGGAGATACTGAGCCACATGGCGTGTGACCTTTTCTCTCTCGTCCGTTCGCTCAAAATTCAGCTCATCTAACTCTTCAAAATAGCCATAGACCTC 2000

CTTGTTTTCTTCTGTTTTGACTCGCGCTATTTTTGTGGCTGCTGAAAGCGGGAAAAATTTAGTATTTTATGAGCTTATCTTTATGCAATACATA 2100

AAAAACGAGGCAATTTAAAAATATTAATAATGAGGTTGTAGATGTAGATTTTGAAGAAAGAAAAAACAAAAAATAGCAACCGCCAGATCAAAA 2200

TTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAATGAAAGTGCCTGCA?AAATCTAGTGTAACTTTAGATTGAACCTGGAA 2300

ATCTAAGCCTGAACATAGCCTTATTCTAGATCTTAGTTGCGCATAGCTCAAGCCCAAGCAGAAATGACTTGCAATTTAGTTAAGCCTAGATTGACTT 2400

GCCTGCTCAGTCTAATCCAGACTAGATTTCCAAGAGATTTTCAATTTTAAATGTTTCCAGTTTCTTGTTACTTAAATCTTAATGCCCTGTGATGCGT 2500

AAAATCGTTATCCCTTTCTCTCACACTTCAATTAAGATTCATCAAGATTTGGTATCAAGCCAAAGAGCTGGAATTAACACCCCTCATCATCAACC 2600

ACTTCATCAATTAATACAAATTCATTCGGTCCGTCGAGCCGTTGAGTGGCAATAAATGTTGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTA 2700

TCCGATCTTCCGGCTCTTTTTAGAAATTAATTTTCAAGATCATCATCAAGCTACAGCTTATTTTGAATCTAAACCGCACTACCTTCCAACTCCA 2800

AAAACTTCTAGACCAACAAACCGCTAGTTGTTGTTGTACAACACAAAAATCGGAAGCTCAAGCTAGCCGCTCCGAAAGCGGTGAGCACCCCAAAA 2900

CTTGCTTCTGTGAAGACTATTGGAGCAAAACAGAGCCGATAACAGCGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3000

donderdag, 27 november 1997 16:46
In 35 PNPB M3p (1 > 12641) Site and Sequence

Page 2

CATCTTCTCATCGAATAGCCACACCTACGAGAAAGCGCGCGCTCAACACAACTTTGTGAAATCGCTGCCCACTGAAAAGTGGCCT 3100
GAAGCGCGGACAGTAAGCTGGGAAGTGCCACGTCTATGTGGAAGCTTTGTACGGTGAGTATTTGAAATCGGAAATGGAAATGTATTTTTTAAAACT 3200
GAAATTTACAAAATAAATTAATAAATTAAGATTTTTCTCTGATAGTATTGCATCCACTATTTTACTTTGAAGATTTATATCTTGTTTCATATTGAAG 3300
ATATCAGATATAGAAAAGAAATAAAATATTTTGACAGTTGATAATTTTGTATAGGACCAAGACAAGTGAGATATAAGCTGTCAAAGTTGATTTTC 3400
AAGAAATTTTAAACCTAGTTTTGCGAAGCTCTGGGCTCATCTATTTAGAACCGATTCTGACTTCTTCGTTCTTCTGACTCTACCAAAACCAAAA 3500
CCAACCTACTAATAAAATGATGAGCAATTGGGAATTTGTCTCCCATTTTCTCTTCTCTTCTGACTCTTTCAGAACTATGTCCCATCTTTTGTG 3600
TTGTGCCCCCATAAAGACTCTTCCGGAATAATGTTGCAACGGAAGTGATATTCGAGCATTTTTCGACGTGAGGGCGGAAAAACACATCTGGCTGA 3700
CAAAGAGTAAGCAATTTCTCAGCTCTTCTCGCGGTTTTCAATTGTTTTCAAAATGAGCTACTACAGAGTGAAGAGCACAAATTGCAAAACATT 3800
TTTGTGTGAGATGCATTTTGAATAATTAACCTTACGTTTTAGTTCTAGTATTTATTTTTCATATAAATTAGAGCTTCTAGACCTCTATATTT 3900
TTTAAACTTCTACTGAAATATACGAGATTCTTTGACTTTCCGGAATTTGTCTTATGGCTCTATTATTTATGAGAAACATTTTTTAAATTTTTT 4000
GAAAAAAACTGTGCATCTCTGTTTTTACATAGTAATTTCCAGCCAAAAGTTTCTACCGTAAACGGACGCCCAATCATATCTCAACAGACTCSA 4100
AACGATGCTCAAAGAGCAGTGAAGAGAGTCCGGATACGCTGGATTCAACAGCAGCTGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCATGCATTC 4200
CACATCTTCCAAGGTTCTGTTTTAGGAGAACTGTTTTTTTGTGTTTCTGACCTTCACATAGTCTCGGATGTTTATAAAGTGAGGTTCTGCGGACAC 4300
CTGCCATAAATGTGAATCCGCCATTTGTTGTCACAAAAAATTTGACAGACCTGCTTTATACATTTTAGGATAAATGTATACGGTATTTGTCAA 4400
ACCCAACTTTTAAATTTTATTTTTCAGATCAAAATTTGATGTTAAAGTTTAAAGATTTTACGAAAAATGTTTACTTAAACTTTTTATATGGATA 4500
AAATTTTAGAATTCAGATAGGAGTTCCGTCCCTAAACCTTTGTTGTCGCCGAGAAGCTTTCGAATTAATATCTCATTTTATAAGTTCGAACTAAT 4600
TTTTTGTGCACTGAATTTTACAGATGAACTTTAAGTTTCAATTTATCCCATTTGAAACCGTCCCTTCTATAAACTTCAAAATTTTACAGATTCAA 4700
CGTCAGACGAAAAGTCTCCGTATCAGACGATCTTACTCTTAACGCTCCATCGTGACAGTATCAGACAGCGGATAGCCGCAACACCGGTTTCTCCAAA 4800
TATTATCAACAGCGCTGTTGAGGTGAGTATTTTTGTTCTGGGTAGAGGCTTCTGTCAAGTTTGGCTAAATTTATTAATTTAGGTTAGAGGCTGGC 4900
AAAGCCATTGATCAAGCATGGGCTAACTGGGCGGCTGAACCATGTACATATCTTTGGCCGAGTGTGCAATCTAAAGATTGGAAGCTGGCTTCA 5000
AAGTCGACTAGGCAAAAGTGCAAAATGGAAATATCTTGAATTTCAATACGTTTCCGTCTTCCATTCTCTTTTTTGTCTCTTTTGTGAGAT 5100
TTTCCCTTTTTAGATTTTAAGATTTTGATACATTTAATGTCTGGCTTCCCTTCTAAGAGCCTTCTATATTTTGAATAAATCAATTTTAGGA 5200
AAAACCAACTGGCAGTGAAAGGAGTGAAAGCACAGCGAAAAAGATCCACCTCCAGCTGTCCGCCAGCTGACACCCAGCCAAATCGGAGTTGTT 5300
AGTCAATTTATGGCATAAGAAGTTGACAAATGGTACGTTTATTTCTGAACCTTACTTATGTTTCCGTCGGTGACGTTTTGTTGACCATGTGATGGGA 5400
AGTAATTTTGGATTATTTAAAGTTTGTGCGGAATAGTAAGGAGGAGTACAATTATTTTGTAGAGGTTTCAAGTAACTTTGATTTTCTGACCA 5500
TAAGTTTTTTCTGAAAGTTGTTTAAAAATTTGAGTTAAAAATTAATAAATCTCTATAAAATTTCTAAATTTCTTGGGAATTTTTTCAAAATGTT 5600
TTTCCAAATATGTCTAATAGTAAGATTTGTTTGTGAATTTACAAACATATTTAAAAATACATTTTAAATTTATTCGATTTTTCTGTTCCGAGGAGTGACA 5700
AAAAATCAGAAAAAGCGAAATTTAATTCAAAAAAATATTTTGAAACTCAGAAATAGAGTACTTTCAAAAAATCAACAAAAAATATCAAAAGAT 5800
TCATATTTTTCAGAAATAGACATAGTCAAAACTATCAAAAAATCTACTATTTTCCGGAATAAATGAGAAAAATCCAAAAATGTAAAAACAAAA 5900
TTGAGAAAAATTCAGAAATTTGAAAAAATCTCTTGTAGGAAAAATTAATAATTTTAAATGTGTGATTTCTGAAACCAAGCATTTTCCGACTTTCCGC 6000

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 3

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTCGCCACGTCAGCAACCGCTTCA
TTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTAAATGGGAGCGGTGCACAGCGGTGCAGTCTTGGCGAAGT
eGFPC.e.unc53xba
C.e.unc53 xba
K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S
GCAACTAACCCAAATTCCAACTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAGATTGGTATCA
CGTTGATTGGGTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTTATAGCTTTAACTAAGTAGTTCTAACCATAGT
eGFPC.e.unc53xba
C.e.unc53 xba
A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I
AGCCAAAGACGCTGGACTTAAACCACCTCATCATCAACCACTTCATCAATAATACAAATTCATTCCGTCGTCGAGCCGTTTCAGTGGCAATAATA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTATTATGTTTAAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTAT
eGFPC.e.unc53xba
C.e.unc53 xba
K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N H
TGTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTGGAATCTAAACCGACCTACCTCCCAACTCCAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTGGCTGGATGGAGGGTTGAGGTTTT
eGFPC.e.unc53xba
C.e.unc53 xba
V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I
CCTTGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTACTTGCTTTAAAAACCTCCCACACCTCCCTCGAA
GGAACCTTAGTGGCTTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTGGAGGGTGTGGAGGGGACT
eGFPC.e.unc53xba
P V D P P D L D N . S . S A I P H L . R F Y L L . K T S H T S P .
CCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAAATAAGCAATAGCATCACAAATTCACAAATAAA
GGACTTTGATTTTACTTACGTTAACAACAACAAATGAACAAATAACGTCGAATATTACCAATGTTTATTTTCGTTATCGTAGTGTAAAGTGTATT
T . N I K . M Q L L L L T C L L Q L I M V T N K A I A S Q I S Q I I
GCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTGTAAGCGTTAATTTTTGTTAAATTCGCGTTA
CGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAACATTGCAATTATAAAACAATTTAAGCGCAAT
H F F H C I L V V V C P N S S M Y L N A . I V S V N I L L K F A L

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 4

AATTTTGTAAATCAGCTCATTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCT
TTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTAGCGGTTTAGGGAATATTAGTTTCTTATCTGGCTCTATCCCAACTCACAACAAS
N F C . I S S F F N Q . A E I G K I P Y K S K E . T E I G L S V V 250

CAGTTTGAACAAGAGTCCACTATTAAGAAGCTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACC
GTCAAACCTTGTCTCAGGTGATAATTTCTTGACCTGAGGTGTCAGTTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGG
P V W N K S P L L K N V D S N V K G R K T V Y Q G D G P L R E P S P 260

CTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTTAAAGGGAGCCCCGATTAGAGCTTGACGGGGAAAGCCGCGAACGTG
GATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGATTTAGCCTTGGGATTTCCCTCGGGGGCTAAATCTCGAAGTGGCCCTTTGCGCCGCTTGAC
S S F L G S R C R K A L N R N P K G S P R F R A . R G K P A N V 270

GCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACACACCCGCGCGCTTAATG
CGCTCTTCTCTCCCTTCTTCTGCTTCTCGCCCGGATCCCGGACCGTTACATCGCCAGTGCGACGCGCATTTGGTGGTGTGGGCGGCGCAATTAC
A R K E G K K A K G A G A R A L A S V A V T L R V T T T P A A L N 280

CGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAATATGTATCCGCTCAT
GCGGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCTTTACACGCGCTTGGGGATAAACAATAAAAGATTATGTAAAGTTATACATAGGCGAGTA
A P L Q G A S G G T F R G N V R G T P I C L F F . I H S N M Y P L M 290

GAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAGGAAGAGTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGA
CTCTGTTATTTGGGACTATTTACGAAGTTATTATACTTTTCTCTCAGGACTCCGCCCTTTCTTGGTGCACACCTTACACACAGTCAATCCACACCTT
R Q . P . . M L Q . Y . K R K S P E A E R T S C G M C V S . G V E 300

AGTCCCAGGCTCCCCAGCAGGAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAGGCTCCCCAGCAGGAGAAG
TCAGGGTCCGAGGGGTCGTCCTTCTATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTCGTCCTCTTC
S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P Q A P Q Q A E 310

TATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGCCAGTTCCGCCATTCTCCGCC
ATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGCGGGGATTGAGGCGGGTCAAGCGGGTAAAGAGGCGGS
V C K A C I S I S Q Q P . S R P . L R P S R P . L R P V P P I L R P 320

CATGGCTGACTAATTTTTTTTATTATGAGAGGCGGAGGCCCTCGGCCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCT
GTACCGACTGATTAATAAATAAATACGTCTCCGGCTCCGGCGGAGCGGAGACTCGATAAGGTCTTCATCACTCTCCGAAAAAACCTCCGGATCCGA
M A D . F F L F M Q R P R P R P L S Y S R S S E E A F L E A . A 330

TTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGAGGTTCTCCGGCCGCTTGGGTGAGAGGCTAT
AAACGTTTCTAGCTAGTTCTCTGTCTTACTCTAGCAAAGCGTACTAAGTTGTTCTACCTAACGTGCGTCCAAGAGGCGGCGAACCACCTCTCCGATA
F A K I D Q E T G . G S F R M I E Q D G L H A G S P A A W V E R L 340

TCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCGTGTTCGGCTGTGAGCGCAGGGGCGCCGGTTCTTTTGTCAAGACCGACCT
AGCCGATACTGACCCGTGTGTCTGTTAGCCGACGAGACTACGGCGGCAAGGCCGACAGTCGCGTCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGA
F G Y D W A Q Q T I G C S D A A V F R L S A Q G R P V L F V K T D L 350

GTCCGGTGGCCTGAATGAAGTGAAGACGAGGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGGCGAGCTGTGCTGACGTTGTCTAGTAA
CAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCGATAGCACCGACCGGTGCTGCGCCGAAGGAACGCGTCAACAGAGCTGCAACAGTACTT
S G A L N E L O D E A A R L S W L A T T G V P C A A V L D V V T E 360

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 5

G G G G A A G G G A C T G G C T G C T A T T G G G C G A A G T G C C G G G C A G G A T C T C C T G T C A T C T C A C C T T G C T C C T G C C G A G A A G T A T C C A T C A T G G C T G A T S C A A
C G C C C T T C C C T G A C C G A C G A T A A C C C G C T T C A C G G C C C G T C C T A G A G G A C A G T A G A G T G G A A C G A G G A C G G C T C T T T C A T A G G T A G T A C C G A C T A C G T T 370
A G R D W L L L G E V P G Q D L L S S H L A P A E K V S I M A D A
T G C G G C G G C T G C A T A C G C T T G A T C C G G C T A C C T G C C C A T T C G A C C A C A A G C G A A A C A T C G C A T C G A G C G A G C A C G T A C T C G G A T G G A A G C C G G C T T T G T 380
A C G C C G C G A C G T A T G C G A A C T A G G C C G A T G G A C G G G T A A G C T G G T G G T T C G C T T T G T A G C G T A G C T C G C T C G T G C A T G A G C C T A C C T T C G G C C A G A A C A
M R R L H T L D P A T C P F D H Q A K H R I E R A R T R M E A G L V 390
C G A T C A G G A T G A T C T G G A C G A A G A G C A T C A G G G G C T G C G C C A G C C G A A C T G T T C G C C A G G C T C A A G G C G A G C A T G C C C G A C G G C G A G G A T C T C G T C G T G
G C T A G T C C T A C T A G A C C T G C T T C T C G T A G T C C C G A G C G C G G T C G G C T T G A C A A G C G G T C C G A G T T C C G C T C G T A C G G G C T G C C G C T C C T A G A G C A S C A C 390
D Q D D L D E E H Q G L A P A E L F A R L K A S M P D G E D L V V
A C C C A T G G C G A T G C C T G C T T G C C G A A T A T C A T G G T G G A A A T G G C C G C T T T T C T G G A T T C A T C G A C T G T G G C C G G C T G G G T G T G G C G G A C C G C T A T C A G G 400
T G G G T A C C G C T A C G G A C G A A C G G C T T A T A G T A C C A C C T T T A C C G G C G A A A G A C C T A A G T A G C T G A C A C C G G C G A C C C A C C C G C C T G G C G A T A G T C C
T H G D A C L P N I M V E N G R F S G F I D C G R L G V A D R Y Q
A C A T A G C G T T G G C T A C C C G T G A T A T T G C T G A A G A G C T T G G C G G C A A T G G G C T G A C C G C T T C C T C G T G C T T T A C G G T A T C G C C G C T C C C G A T T C G C A G C G 410
T G T A T C G C A A C C G A T G G G C A C T A T A A C G A C T T C T C G A A C C G C G C T T A C C C G A C T G G C G A A G G A G C A C G A A T G C C A T A G C G G C G A G G G C T A A G C G T C G C
D I A L A T R D I A E E L G G E W A D R F L V L Y G I A A P D S Q R
C A T C G C C T T C T A T C G C C T T C T T G A C G A G T T C T T C T G A G C G G G A C T C T G G G G T T C G A A T G A C C G A C C A A G C G A C G C C C A A C C T G C C A T C A C G A G A T T T C G 420
G T A G C G G A A G A T A G C G G A A G A A C T G C T C A A G A A G A C T C G C C T G A G A C C C A A G C T T T A C T G G C T G G T T C G C T G C G G G T T G G A C G G T A G T G C T C T A A A G C
I A F Y R L L D E F F . A G L W G S K . P T K R R P T C H H E I S
A T T C A C C G C C G C C T T C T A T G A A A G G T T G G G C T T C G G A A T C G T T T T C C G G G A C C C G G C T G G A T G A T C C T C A G C G C G G G A T C T A T G C T G G A G T T C T T 430
T A A G G T G G C G G C G G A A G A T A C T T T C C A A C C G A A G C C T T A G C A A A G G C C C T G C G G C G A C C T A C T A G G A G G T C G C G C C C T A G A T A C G A C C T C A A G A A
I P P P P S M K G W A S E S F S G T P A G . S S S A G I S C V S S
C G C C C A C C C T A G G G G A G G C T A A C T G A A A C A C G G A A G G A G A C A A T A C C G G A A G G A A C C C G C G C T A T G A C G G C A A T A A A A G A C A G A A T A A A A C G C A C G G T 440
G C G G G T G G G A T C C C C T C C G A T T G A C T T T G T G C C T T C C T G T T A T G G C C T T C C T T G G G C G C A T A C T G C C G T A T T T T T C T G C T T A T T T T G C G T G C C A
S P T L G G G . L K H G R R O Y R K E P A L . R Q . K D R I K R T V
G T T G G G T C G T T T G T T C A T A A C G C G G G T T C G G T C C C A G G G C T G G C A C T C T G T C G A T A C C C A C C G A G A C C C A T T G G G G C C A A T A C G C C C G C G T T T C T T 450
C A A C C C A G C A A A C A A G T A T T T G C G C C C A A G C C A G G G T C C C G A C C G T G A G A C A G C T A T G G G G T G G C T C T G G G G T A A C C C C G G T T A T G C G G G C G A A A G A A
L G R L F I N A G F G P R A G T L S I P H R D P I G A N T P A F L
C C T T T T C C C A C C C A C C C C C A A G T T C G G G T G A A G G C C A G G G C T C G C A G C C A A C G T C G G G G C G C A G G C C C T G C C A T A G C C T C A G G T T A C T A T A T 460
G G A A A A G G G G T G G G G T G G G G G T T C A A G C C C A C T T C C G G G T C C C G A G C G T C G G T T G A G C C C C C C G C T C G G G A C G G T A T C G G A G T C C A A T G A G T A T A T A
P F P H P T P Q V R V K A Q G S Q P T S G R O A L P . P Q V T H I
A C T T T A G A T T G A T T T A A A C T T C A T T T T A A T T T A A A A G G A T C T A G G T G A A S A T C C T T T T T G A T A T C T C A T G A C C A A A T C C C T T A A C G T G A G T T T T C G 470
T G A A A T C T A A C T A A A T T T T G A A G T A A A A T T A A A T T T C C T A G A T C C A C T T C T A G G A A A A C T A T T A G A G T A C T G G T T T T A G G G A A T T G C A C T C A A A A G C
Y F R L I . N F I F N L K G S R . R S F L I I S . P K S L N V S F R
T T C C A C T G A G C G T C A G A C C C G T A G A A A A G A T C A A A G G A T C T T C T T G A G A T C C T T T T T T C T G C G C G T A A T C T G C T G C T T G C A A A A A A A A C C A C G G C 480
A A G G T G A C T C G C A S T C T G G G C A C T T T T T C T A G T T C C T A G A A G A A C T C T A G G A A A A A A G A C G C G A T T A G A C G A C G A A C G T T T G T T T T T T G G T T G G G
S T E R O T P . K R S K D L L E I L F F C A . S A A C K Q K N H R

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 6

TACCAGCGGTGGTTTGGTTTCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGCAGATACCAAACTGTCCTTCTAGT
ATGGTCGCCACCAACAAACGGCCTAGTTCGATGGTTGAGAAAAAGGCTTCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCA 530
Y Q R V F V C R I K S Y Q L F F R R . L A S A E R R Y Q I L S F .
GTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC GCCTACATACCTCGCTCTGCTAATCCTGT TACCAGTGGCTGCTGCCAGTGGCGATAAG
CATCGGCATCAATCCGGTGGTGAAGTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATT 540
C S R S . A T T S R T L . H R L H T S L C . S C Y Q V L L P V A I S
TCGTGCTTACC GGGTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAA
AGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTCGGGTGGAACCTCGCTT 550
R V L P G W T O D D S Y R I R R S G R A E R G V R A H S P A W S E .
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGT
GCTGGATGTGGCTTGACTCTATGGATGTGCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCGCGCTGTCCATAGGCCATTCCCGTCCCA 560
R P T P N . D T Y S V S Y E K A P R F P K G E R R T G I R . A A G
CGGAACAGGAGAGCGCAGAGGGAGCTTCCAGGGGGAACGCCTGGTATCTTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTG
GCCTTGCTCTCGCGTGTCCCTCGAAGGTCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAC 570
S E Q E S A R G S F Q G E T P G I F I V L S G F A T S D L S V D F C
TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCTTTTACGGTTCTTGGCCTTTTGTGGCCTTTTGTGCTCACATGTCTTTT
ACTACGAGCAGTCCCCCGCTCGGATACCTTTTTCGGTCTGTTGCGCGGAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTACAAGAAAG 580
D A R Q G G G A Y G K T P A T R P F Y G S W P F A G L L L T C S F
CTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCATGCAT 5447
GACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L R Y P L I L V I T V L P P C I

Tuesday, 18 November 1997 11:48

fig 34 pLM4 (1 > 10070) Site and Sequence

Enzymes : 100 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

100
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGC
pCMV
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
200
CCCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAAGTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTCATAAATGCCA
pCMV
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
300
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCT
pCMV
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
400
CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACACGCAAAACCGTCATGTAGTTACCCGCACCT
pCMV
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q W A V
500
TAGCGGTTGACTCACGGGGATTTCGAAGTCTCCACCCATTGACGTCAATGGGAGTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAATGAGTGCCCTAAAGGTTCAAGGTGGGGTAAGTGCAGTTACCTCAACAAAACCGTGGTTTAGTTGCCCTGAAAGGTTTTACAGCAT
pCMV
I A V . L T G I S K S P P H . R Q W E F V L A P K S T G L S K M S
600
ACAAC TCCGCCCCATTGACGCAATGGGCGGTAGGCGGTACGGTGGGAGGCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAAGTGCCTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTCGCTCGACCAATCACTTGGCAGTCTAGCGGATCGCGAT
pCMV
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
700
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACACGGGTAGGACAGCTCGACCTGCCCGCTGCATTGCCGGTGTCAAGTCGG
EGFP
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
800
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTATCTGCACACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGTGGGAGCACTG
EGFP
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
900
CAECC TGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAG
GTGGGAC TGGATGCCGCACGTACGAAGTGGCGATGGGGCTGGTGTACTTCGTCGTCTGAAGAAGTTACGGCGGTACGGGCTTCGATGCAGGTCTC
EGFP
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CGCACCATCTTCTTCAAGGACGACGGCAAC TACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCTGCTGCCGTGTATGTTCTGGGCGCGGCTCCACTTCAAGCTCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC

EGFP

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGCTCTATATCATGGCCGACAAGCAGAAGAAGCGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCGAGATATAGTACCGGCTGTTCTGCTTCTTGGCGTAGTT

EGFP

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I

GGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTG
CCACTTGAAGTTCCTAGGCGGTGTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGTATGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC

EGFP

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCATATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGA
GGGCTGTGGTGTATGGACTCGTGGGTGACGGCGGACTCGTTTCTGGGGTGTCTCTTCGCGCTAGTGTACAGGACGACCTCAAGCACTGGCGGGCGGCCCT

EGFP

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCGACGTCGATAAGCTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTTCAAGCTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTCTGAATATAGCTTAAGGACGTCGG

EGFP

I T L G M D E L Y K S G L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGCTCTCTTCTCTGTCACGCTTACAGGACCTAGAGGCCGCTTGGACCTCTCTGGTACAGGTCG

insert pLM1

ORF pLM1

L L F S Q M L D P E S Q R K R T V Q N V L D L R Q N L E E T M S S

CTGCGAGGGTCCCAGGTGACTCACAGTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGACGGTGTCCAGCCTCTCCAACCGCTCGT
GACGCTCCCAGGGTCCACTGAGTGTGAGGGACCTCTACTGGACGATGCTGTCGCTACTACGGTGGGTGCGTCCGACAGGTGAGAGGTTGGCGAGCA

insert pLM1

ORF pLM1

L R G S O V T H S S L E M T C Y D S D D A N P R S V S S L S N R S

CCCTCTGTCTATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCTG
GGGAGACAGTACCGGATACCGGTGAGGTGAGGCGCGGACGTCGACCACTGCGCGGGAGACACCCACCTCGACGGCGAGCTCCCTTGGGGCGGAC

insert pLM1

ORF pLM1

S P L S W R Y G Q S S P R L Q A G D A P S V G G S C R S E G T P A W

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 3

GTACATGCACGGCGAACGGGCCCCTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCCTGGAGCTGGTCGAATCCCTG
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGGTGTGGTACGGGTACGCGTCGGGGTCGTTTCGAGTCGGTATAGAGGGCGGACCTCGACCAAGCTTAGGGAC

insert pLM1

ORF pLM1

Y M H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L

GACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG
CTGAGCC TACTCCACCTGGAGTTCAGGCCGATGACTCGCTGTCACTGGAGTACCGTTCGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCCTAC

insert pLM1

ORF pLM1

D S D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D

AAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTTCAAGAAGTTCATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCC
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGG

insert pLM1

ORF pLM1

E S S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P

CACTGCTTCTCGCAGGAACCTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTAGTGAATCAGAGGAG
GTGACGAAGAGCGTCTTGAGTTGTTATCAGATGCGTGCTGAGTCTCTCGCGAGTGACCGTCTTTCACCGACTCGACCAAACTCACTTAGTCTCTC

insert pLM1

ORF pLM1

T A S R R N S T I V L R T D S E K R S L A E S G L S W F S E S E E

AAAGCCCCATAAAACCTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCAT
TTTCGGGGATTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCTGAAGATTACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTA

insert pLM1

ORF pLM1

K A P K K L E Y D S G S L K M E P G T S K W R R E R P E S C D D S

CCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCCTGGTTCCCTGAAGAAGGCGCAAGACCCACCTGTGGCTGTAAC TTCCCCATCACTCA
GGTTCCACCTCTTGACTTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGSACTTCTTCCCGTTCGGGGTGGACACCGACATTGAAGGGGGTAGTGGT

insert pLM1

ORF pLM1

S K G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

CACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCC
GTGTGCGGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCTTCGAACGTCACCTTCTATGACCCGAGGTTGCGAGGAGG

insert pLM1

ORF pLM1

T A Q S A L K V A G K P E G K A T D K G K L A V K N T G L O R S S

TCTGATGCTGGTGGGACCGCCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCTCC
AGACTACGACCGCCCTGGCGGACTCAC TACGATTCTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGAGGAG

insert pLM1

ORF pLM1

S D A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P

CTGCCACAGGCACAGCCACTGTCATGCAAAGTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCG
GACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTACCCGCG

insert pLM1

ORF pLM1

P A T G T A T V M Q T G G S A T L S K I O K S S G I P V K P V N G R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCTTGCTCTCGGAGCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGCCAGCCAAAG
GTTCTGATCGAATCTACAAGGTGTGCGCTCTCGGTCTTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGECGGTCTGCTC

insert pLM1

ORF pLM1

K T S L D V S N S A E P G F L A P G A R S N I O Y R S L P R P A I

TCAAGTTCTATGAGCGTGACCGGGGGGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCTCAGCACCAAGCAGGAGGCCCTTACGC
AGTTCAAGATATCGCACTGGCCGCCGCCGCCACCTGGAGCGGGACACTCGTCTGCTGAAGTGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCG

insert pLM1

ORF pLM1

S S S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T

CTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGCGGACCACTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGC
GAAGGTCGACTTCTCGGATGGTTCATCGGTACCCGCCCTGGTGAGGTGCGGGACAGTGTCTGTCTAGCCCTTTCTCTTCGGGTTTCGGTTCGG

insert pLM1

ORF pLM1

P S R L K E P T K V A S G R T T P A P V N Q T D R E K E K A A A A

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

AGTGGCCTTGGA CTGAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCA 300
TCACCGGAACCTGAGTCTGTTGTAGAGGAACCTTCATAACCGAGGGGTCTCTCATGAGGGTCTTGGTTCGTTCCGTTGGGGTGTCCGTTGGTTCGACCGT
-----insert pLM1-----
-----ORF pLM1-----
V A L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A
GAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCAACCTCAGTACCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTAC 310
CTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAAGTGTCCAGTTGAGGTGTGACAGCTAGATG
-----insert pLM1-----
-----ORF pLM1-----
E L P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L
CATCATCCAGTGATACCAACCATGCTTCAAAGGTCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCTGCTTACCCCCAGTCC 320
GTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTTCGAGTCGTAGACCCCGGGAGAGGGAAGGACGAAGTGGGGTTCAGG
-----insert pLM1-----
-----ORF pLM1-----
P S S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P
GGCACCCATCCTCAATATTAACCTCAGCCAGCTTCTCCAGGGCC TGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGATGTACCCCAAAC TC 330
CCGTGGGTAGGAGTTATAATTGAGTCGGTCAAGAGGGTCCCGGACCTCGATTACTACCAAAGTCACACGGTTTCTCTGGGCGTACATGGGGTTTGAG
-----insert pLM1-----
-----ORF pLM1-----
A P I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L
TCAGGCC TGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTG 340
AGTCCGGACGTGCTCTGTACCTCAGGGAGGTCTACGGT TAC TCGGAGGGGTACGGAAGGGTCTCATGAGGGCAGGGGTGGGGTGGACGAGGGGAC
-----insert pLM1-----
-----ORF pLM1-----
S G L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P
CTGTCCCCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAA 350
GACGAGGGGTGCTTCTTCTCTGCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCTATTAGTCGCCCTAGCCTTGTGAGAAGGGTT
-----insert pLM1-----
-----ORF pLM1-----
A A P T E E E T E E L T V S G S P R A G Q L D S N Q R D R N T L P P

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

GAAAGGGCTCAGGTACCACTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATCCCATACCATGGTGGGCTGCTGAATCCGATGACCAGTCAGAG
CTTCCCCGAGTCCATGGTCGAAGTCAGGGTCTCTCTGGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTC

insert pLM1

ORF pLM1

K G L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E
CTGCCTTCTCCCCCTGCACCTCCCATGTCTCTGAGTGCAAAGGGCCAACCTACCAACATAGTGAGTCCACGCGGCCACCAAGCAATCACCCTG
GACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCCTTCTAGTGGGCGA

insert pLM1

ORF pLM1

L P S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R
CCAACAGCATCCCCACCCACGAGGCGGCTTCGAGCTGTACAGCGGCTCCCAATGGGAGCACCTGTCCCTGGCCGAGAGACCAAGGAATGATTG
GGTTGCTGAGGGTGGGTGCTCCGCCGAAGCTCGACATGTGCGCGAGGGTTACCCCTCGTGGGACAGGACCGGCTCTCTGGGTTCCCTTACTAAGC

insert pLM1

ORF pLM1

S N S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R
GTCAGGATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGCTGTCCCTGGCTCCAGTGCCTCCTCCACTACTCCTCAGCTGAGGAGAGGATG
CAGTCTAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTCACGACAGGGACCGAGGTACCGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTAC

insert pLM1

ORF pLM1

S G S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M
CAATCTGAGCAAATCCGAAGCTTCGTAGGGAAC TGAATCATCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTCTGCCAATGCTAATCTGGTGG
GTTAGACTCGTTTAGGCTTCGAAGCATCCCTTGACCTTAGTAGGGTCTTTTACCCTGGGAAC TGCAGAGTCGAAGACGGTTACGATTAGACCAAC

insert pLM1

ORF pLM1

D S E O I R K L R R E L E S S Q E K V A T L T S Q L S A N A N L V
CTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCTGCGACACC TGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTGCGAGAAAC
GACGAAAAC TCGTCTCGGACCATTATACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGCTCTCTTCTGTGACTCGACGACCTAAACGCTCTTTG

insert pLM1

ORF pLM1

A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T

Tuesday, 18 November 1997 13:56
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 3

AGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACTTGGATCA
TCTTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGT

U3 stuk

ORF

K K E V S E L R S E L V E K E M K L T D I R L E A L N S A H Q L D Q

GCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACT
CGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGA

U3 stuk

ORF

L R E T M H N M Q L E V D L L K A E N D R L K V A P G P S S G S T

CCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCACGCCCTCCCTAGGCTGGCACTCACCCATTCTTCGGCCCCAGTCTTGCAGACACAGACC
GGTCCCTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGCTGTGTCTGG

U3 stuk

ORF

P G Q V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D

TGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCCCGAGCACATCATCAAGGGGACTT
ACAGTGGGTACCTACCGTAGTCATGAACACAGGTTTCTCTTCACTGGGAGGCCCCACCACCACTCCTACGGGGGCGTCGTGTAGTAGTTTCCCTTGAA

U3 stuk

ORF

L S P M D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L

GAAGCAGCAGGAATCTTCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTCCAAGTGTTCAGGACTATATT
CTTCGTCGTCTTAAGAAGSACCCGACATCGTTCAGTCACTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCCTGATATAA

U3 stuk

ORF

K Q Q E F F L G C S K V S G K V D W K M L D E A V F Q V F K D Y I

TCTAAATGGACCCAGCCTCTACCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAACAGAGTGTGGATGCAGAGCCCCCG
AGATTTTACCTGGGTCGGAGATGGGACCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTGTCTCACAACCTACGTCTCGGGGGG

U3 stuk

ORF

S K M D P A S T L G L S T E S I H G Y S I S H V K R V L D A E P P

Tuesday, 18 November 1997 13:57
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 4

AGATGCCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCC
TCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTCCAGACTTCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGCT 3000

U3 stuk

ORF

E M P P C R R G V N N I S V S L K G L K E K C V D S L V F E T L I P

CAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACC TACCTGACCAATCGG
GTTTCGGCTACTACGTCGTGATGTATTTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCG 3100

U3 stuk

ORF

K P M M O H Y I S L L L K H R R L V L S G P S G T G K T Y L T N R

TTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGT
AACCGGCTCATGGACCACCTCGCGAGACCAGGCACTCCAGTGTCTCCCGTAGCAGTCTGGAAGTTGTACGTGGTCTGTCAGAACGTTCTTAGACGTTGACA 3200

U3 stuk

ORF

L A E Y L V E R S G R E V T E G I V S T F N M H Q Q S C K D L Q L

ATCTTTCCAACCTAGCCAACAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGC TCCATCAG
TAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCC TACACGGGACCAC TAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTC 3300

U3 stuk

ORF

Y L S N L A N O I D R E T G I G D V P L V I L L D D L S E A G S I S

TGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTG
ACTCAACCAAGTTACCCCGGGAGTGGACSTTCATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAAC 3400

U3 stuk

ORF

E L V N G A L T C K Y H K C P Y I I G T T N Q P V K M T P N H G L

CAC TTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCG
GTGAAC TCGAAGTCTACAAC TGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGC 3500

U3 stuk

ORF

H L S F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S

Tuesday, 18 November 1997 13:57
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 6

ACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGA
TGTAGTTACGGTGTTCCTTCTCGACGAAGCCACGAGCTGACCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTC 360

U3 stuk

ORF

D I N A N K E E L L R V L D V V P K L V Y H L H T F L E K H S T S D

CTTCTCATCGGCCCTTGCTTCTTCTGTCGTGTCCATTGGCATTGAGGACTTCGGACCTGGTTCATTGACCTGTGGAACAACCTATCATTCCCTAT
GAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCC TGGACCAAGTAAC TGGACACCTTGTGAGATAGTAAGGGATA 370

U3 stuk

ORF

F L I G P C F F L S C P I G I E D F R T V F I D L W N N S I I P Y

CTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCTGGCCATCAG
GATGTCTTCTCGGTTCCTACCTATTTCAGGTACCTGCTTCGACGAACCTCTGGGTACCTTACCCAGGCCCTGTGTGAAGGACCGGTAGTC 380

U3 stuk

ORF

L Q E G A K D G I K V H G Q K A A V E D P V E W V R D T L P V P S

CCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCAC
GGGTGTCTGTTAGTTTCGACATGGTGGACGGGGSTGGGTGGCACCCGGGAGTGTCTAACGGAGTGGAGGGCTCCTATCTGTCAAGTTCTGTCTGTS 390

U3 stuk

ORF

A Q Q D O S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T

CCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAATACATTGAGTCTCCAGATCAGAAACCATCTGGAC
GGGTTCAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTCTTCGACGGTGTATGTAAC TCAGAGGTC TAGCTCTTTGGTAGGACCTG 400

U3 stuk

ORF

P S S L D S D P L M A M L L K L Q E A A N Y I E S P D R E T I L D

CCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACGTGACCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCTCTCCCTCTCC
GGGTGGAAGTCCGTTGTGAATTCCTAAGCCGTTAGTGACAGTGGGGGCTGTCTGCTTTCGACCGTAGTCGATAGAATCGAAGGAGAGGGGAGAGG 410

U3 stuk

ORF

P N L Q A T L . G F G N H C H P R T A E R V H O L S . L L L S P L

Tuesday, 18 November 1997 13:57
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 6

TCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGAGATGAAAGAGGAGGGACAGGTTCTTGGTGC GTACCTTTGAGAACTT
AGAAAGTCTCGTGACCGAGAGGTGCGGGTCCTCCTCTTGCTCCCTCCTCTACTTTCTCCCTCCCTGTCCAAGAACCACGACATGGAACTCTTGAA 420

U3 stuk

L F Q S T G S P A P G G E Q E G G G D E R G G T G S V C C T F E N F

CCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAAC TTGTGCCCTTAACACATTTACTGGCCCTCTAGAGCGGCCGCCACCGGGTGGAGCTCCAATT
GGATCCTTCCTTACCACCCACCGCAAACCTTGAACACGGGGATTGTGTAAATGACCGGAGGAGATCTCGCCGGCGGTGGCGCCACCTCGAGGTTAA 430

U3 stuk

L G R N G G V A F G N L C P L N T F T G L L . S G R H R G G A P I

CGCCCTATAGTGAGTCGATTACGCGCGCTCACTGGCGTCGTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATCGCCTTGAGC
GCGGATATCACTCAGCATAATGCGCGGAGTGACCGGCAGCAAATGTTGCAGCACTGACCCTTTGGGACCGCAATGGGTGAATTAGCGGAACGTG 440
R P I V S R I T R A H V P S F Y N V V T G K T L A L P N L I A L Q

ACATCCCCCTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGT
TGAGGGGAAAGCGGTGACGCGCATTATCGCTTCTCGGGCGTGGCTAGCGGGAAGGGTTGTCAACGCGTCGGACTTACCGCTTACCTGCGCGGGACA 450
H I P L S P A G V I A K R P A P I A L P N S C A A . M A N G T R P V

AGCGGCGCATTAAAGCGGGGGGTGGTGGTTACGCGCAGCGTGACCGC TACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTCCCTTCCT
TCGCCGCGTAATTCGCCCGGCCACACCACCAATGCGCGTGCCTGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGGAAGGA 460
A A H . A R R V V W L R A A . P L H L P A P . R P L L S L S S L P

TTCTGCCACGTTTCGCCGGCTTTCCCCGTC AAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCAAAAAACT
AAGAGCGGTGCAAGCGGCCGAAAGGGGAGTTTCGAGATTTAGCCCCGAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGA 470
F S P R S P A F P V K L . I G G S L . G S D L V L Y G T S T P K N

TGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTT
ACTAATCCCACTACCAAGTGCATACCCGGTAGCGGGACTATCTGCCAAAAGCGGGAAC TGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACA 480
L I R V M V H V V G H R P D R R F F A L . R V S P R S L I V D S C S

CAAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTTGATTATAAGGGATTTTCCGATTTCCGGCTATTGGTTAAAAAATGAGCTGATTTAAC
GTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATCCCTAAAACGGCTAAAGCCGGATAACCAATTTTTACTCGACTAAATTG 490
K L E Q H S T L S R S I L L I Y K G F C R F R P I G . K M S . F H

AAAAATTTAACGGAATTTTAAACAAAATATTAACGCTTACAATTTAG 4947
TTTTTAAATGCGCTTAAATTTGTTTATAATTGCGAATGTTAAATC
K N L T R I L T K Y . R L Q F R

Tuesday, 18 November 1997 13:57

fig 54 pl.M1 (1 > 8285) Site and Sequence

Enzymes : 72 of 148 enzymes (Filtered)

Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

GTGGCACTTTTCGGGAAATGTGCGGGAACCCCTATTGTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCGATAAAT
100
CACCGTGAAAAGCCCTTTACACGCGCTTGGGGATAACAAATAAAAGATTATGTAAGTTTATACATAGCGGAGTACTCTGTTATTTGGGACTATTTA
G G T F R G N V R G T P I C L F F . I H S N M Y P L M R Q . P . . M

GCTCAATAATATTGAAAAGGAGATGAGTATTCAACATTTCCGTGTCGCCCTTATCCCTTTTTCGGGCATTTGCCTTCCTGTTTGTCTCAC
200
CGAAGTTATTATACTTTTCTCTCTACTCATAGTTGTAAGGCACAGCGGAATAAGGAAAAAACGCCGTAACCGGAAGGACAAAAACGAGTG
L Q . Y . K R K S M S I Q H F R V A L I P F F A A F C L P V F A H

CCAGAAACGCTGGTGAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAACGATCTCAACAGCGGTAAAGATCCTTGAGAGTT
300
GGCTTTTGGCACCCTTTCATTTTCTACGACTTCTAGTCAACCCACGTCCTCACCCTAATGTAAGTACCTAGAGTTGTCGCCATCTAGGAACCTCAA
P E T L V K V K D A E D Q L G A R V G Y I E L D L N S G K I L E S

TTCCGCCCCGAAGAACGTTTCCAATGATGAGCACTTTTAAAGTTCGTATGTGGCGCGGTATTATCCGATTGACGCCGGCAAGAGCAACTCGGTG
400
AAGCGGGCTTCTTGAAGGTTACTACTCGTGAAGTTTCAAGACGATACCGCGCCATAAGGGCATAACTCGGCCCGTTCTCGTTGAGCCAGC
F R P E E R F P M M S T F K V L L C G A V L S R I D A G O E Q L G R

CCGCATACACTTCTCAGAATGACTTGGTGTAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCACTGCTGCC
500
GGCGTATGTAGATAGAGTCTTACTGAACCACTCATGAGTGGTCAGTGTCTTTCTGTAAGTGCCTACCGTACTGTCTTCTTAATACGTCACGACGG
R I H Y S Q N D L V E Y S P V T E K H L T O G M T V R E L C S A A

ATAACCATGAGTGATAACACTGCGGCCAATTAATCTTGACAAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTCGCAACATGGGGGATCATGTAA
600
TATTGGTACTCACTATTGTGACGCGGTTGAATGAAGACTGTGTAGCTTCTCGCTTCTCGATTGGCGAAAAACGTTGTACCCCTAGTACATT
I T M S D N T A A N L L L T T I G G P K E L T A F L H N M G D H V

CTCGCTTGTGCTGTTGGGAACCGGAGCTGAATGAAGCATACCAACGACGAGCGTGACACACGATGCCGTAGCAATGGCAACAACGTTGCGCAAACT
700
GAGCGGAATAGCAACCTTGGCTTCGACTTACTTCGGTATGGTTTGTGCTGCGACTGTGGTGTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA
T R L O R V E P E L N E A I P N D E R D T T M P V A M A T T L R K L

ATTAACGTGGCACTACTTACTCTAGCTTCCCGCAACAATTAATAGACTGGATGGAGGCGGATAAGTTGCGAGCACCTTCTGCGCTCGGCCCTTCCG
800
TAATGACCGCTTGATGAATGAGATCGAAGGCGGTTGTAAATATCTGACCTACCTCCGCTAATTCACGTCCTGGTGAAGACGCGAGCGGGAAGGC
L T G E L L T L A S R O O L I D V M E A D K V A G P L L R S A L P

GTGGCTGGTTTATTGCTGATAAATCGGAGCCGGTGGGCTCGCGGTATCATTCGACGACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
900
CGACGACCAATAACGACTATTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGCTGACCCGGCTTACCATTGGGAGGGCATAGCATC
A G V F I A D K S G A G E R G S R G I I A A L G P D G K P S R I V

TTATCTACACGCGGGGAGTCAGGCAATATGGATGAACGAAATAGACAGATCGCTGAGATAAGTGCCTCATGATTAAAGCATTGGTAACGTCAGACCA
1000
AATAGATGTGCTGCCCCGAGTCCGTGATACC TACTTGCTTTATCTGCTAGCGACTCTATCAGCGAGTGACTAATTCGTAACCATGACAGTCTGGT
V I Y T T G S O A T M D E R N R O I A E I G A S L I K H V . L S D O

AGTTTACTCATATATACTTTAGATTGATTTAAACTTCAATTTAATTTAAAGGATCTAGGTAAGATCCTTTTGTAAATCTCATGACCAAAATCCCT
1100
TCAATGAGTATATGAAATCACTAAATTTGAAGTAAATTAATTTTCTAGATCCACTCTAGGAAAAACTATTAGAGTACTGGTTTAGGGA
V Y S Y I L . I D L K L H F . F K R I . V K I L F D N L M T K I P

TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAAGAAAGATCAAGGATCTTCTTGAGATCCTTTTTCGCGCGTAATCTGCTGCTTGCAA
1200
ATTGCACTCAAAAGCAAGGTGACTGCGAGTCTGGGGCATCTTTCTAGTTTCTAGAGAATCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTT
R E F S F H . A S D P V E K I K G S S . D P F F L R V I C C L Q

CAAAAAACCCGCTACCGCGGTGGTTGTTTGGCGGATCAAGAGTACCAACTCTTTTTCGAAGGTAAGTGGCTTACGAGAGCGCAGATACCAAA
1300
GTTTTTTGGTGGCGATGGTGCACCAACAAACGCGCTAGTCTCGATGGTTGAGAAAAAAGTCTTCAATTGACCGAAGTCTGCTCGCGTCTATGGTTT
T K K P P L P A V V C L P D Q E L P T L F P K V T G F S R A Q I P N

TACTGCTCTTCTAGTGTAGCGGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAGCGCTAATACCTCGCTCTGCTAATCCTGTTACCACTGGCTGCT
1400
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATGTTGGCGGATATGGAGCGAGACATTAGGACAATGGTCAACGACGA
T V L L V . P . L G H H F K N S V A P P T Y L A L L I L L P V A A

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 2

GCCAGTGGCGATAAGTCGTGCTTACCGGGTGGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGC
CGGTCAACCGCTATTTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCTATTTCGCGTCGCCAGCCGACTTCCCCCAAGCACGTGTGTCG 1500
A S G D K S C L T G L D S R R . L P D K A Q R S G . Y G G S C T Q
CCAGCTTGGAGCGAACGACCTACACCGAAGTGAATACCTACAGCGTGAGCTATGAGAAAGCGCCAGCTTCCCGAAGGGAGAAAGCGGACAGGTATCC
GGTCGAACCTTCGCTTGGATGTGGCTTGACTCTATGGATGTGCGACTCGATACCTTTCCGGTGGCAAGGGCTTCCCTCTTCCGCTGTCCATAGG 1600
P S L E R T T Y T E L R Y L O R E L . E S A T L P E G R K A D R Y P
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTT
CCATTCGCGTCCAGCCTTGTCTCTCGCGTGTCCCTCGAAGGTCGCCCTTTCGGGACCATAGAAATATCAGGACAGCCAAAGCGGTGGAGACTGAA 1700
V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L . L
GAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCCTGGCCTTTTGTGCGCTTTTG
CTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTTCGGTGTGCGCGGAAAAATGCCAAGGACCGAAAAACGCGGAAAAAC 1800
E R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F
CTCACATGTTCTTTCCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCGCAACGACCGAGCG
GAGTGACAGAAGGAGCGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGAAACTCACTCGACTATGGCGAGCGCGCTCGGCTTGTGCTCGC 1900
A H M F F P A L S P D S V D N R I T A F E . A D T A R R S R T T E R
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAACCGCCTCTCCCGCGGCTTGGCGGATTCAATATGACGCTGGCAGCAGAGGTT
GTCGCTCAGTCAGTCGCTCTTCGCTTCTCGCGGTTATCGGTTTGGCGGAGAGGGGCGCGCAACCGCTAAGTAATTACGTCGACCGTGTGTCAAA 2000
S E S V S E E A E E R P I R K P P L P A R V P I H . C S V H D R F
CCCGACTGGAAGCGGGCAGTGAGCGCAACGAATTAATGTGAGTTAGCTCACTATTAGGCAACCGAGGCTTACACTTTATGCTTCCGGCTCGTATGT
GGGC TGACCTTTCCCGCGTCACTCGGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGTCCGAAATGTGAAATACGAAGCGGAGCATACA 2100
P D V K A G S E R N A I N V S . L T H . A P Q A L H F H L P A R H
TGTGTGAATGTGAGCGGATAACAATTTACACAGGAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAGCT
ACACACCTTAACACTCGCTATTGTAAAGTGTGCTTTGTGATAGTGGTACTAATGCGGTTCGCGGTTAATTGGGAGTGATTCCCTTGTTTTGA 2200
L C G I V S G . Q F H T G N S Y D H D Y A K R A I N P H . R E Q K L
GGGTACCGGGCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAG
CCATAGGCGGGGGGAGCTCCAGCTGCCATAGCTATTGAACTATAGCTTAAGGACGTCGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTCT 2300
insert pLM1
ORF pLM1
G T G P P L E V D G I D K L D I E F L Q P L L F S Q M L D P E S Q
AGAAAGAGGACAGTCGAGAATGTCTGGATCTCCGCGAGAACCCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTACAGCTCCCTGGAGA 2400
TCTTCTCCTGTACGCTTACAGGACCTAGAGGCGCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCAGGGTCCACTGAGTGTGAGGGACCTCT
insert pLM1
ORF pLM1
R K R T V Q N V L D L R Q N L E E T M S S L R G S O V T H S S L E
TGACCTGCTACGACAGCGATGATGCCAACCCACGCGAGGTGTCAGCCTCTCAACCGCTCGTCCCCTCTGTATGGCGCTATGGCCAGTCCAGTCCGG 2500
ACTGGACGATGCTGCTGCTACTACGGTTGGGTGCGTGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAGACAGTACCGGATACCGGTGAGGTGAGGCGC
insert pLM1
ORF pLM1
H T C Y D S D D A N P R S V S S L S N R S S P L S V R Y G S S S P R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 3

CTGTCAGGCTGGTGACGCGCCCTCTGTGGG TGGGAGCTGCGCGCTCGGAGGGGACGCGCCCTGGTACATGCACGGCGAACGGGCCCTACTCCCACACC 2600
CGACGTCCGACCAC TGCGCGGGAGACACCCACCTCGACGGCGAGCCTCCCTTGC GGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGG
----- insert pLM1 -----
----- ORF pLM1 -----
L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H T
ATGCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCGAA TCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGA 2700
TACGGGTACGCGTCGGGGTCGTTTCGAGTCGGTATAGAGGGCGGACCTCGACCAAGCTTAGGGACCTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACT
----- insert pLM1 -----
----- ORF pLM1 -----
M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M
GCGACAGTGACCTCATGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGACTCAGCGATGCCTC 2800
CGCTGTCACTGGAGTACCGGTTCTGCTACTGCTTCTACTGTAGTGATGGCCGACCTACTTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAG
----- insert pLM1 -----
----- ORF pLM1 -----
S D S D L H G K Y M T E D D D I T T G V D E S S S I S S G L S D A S
AGACAATCTCAGTTCAGAAGAATCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAACCAACAATAGTGCTACGC 2900
TCTGTTAGAGTCAAGTCTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGGAGGTTTCATGAGGGTGACGAAGAGCGTCTTGAGTTGTTATCACGATGCG
----- insert pLM1 -----
----- ORF pLM1 -----
D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L R
ACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCTAAAAAAGTGGAGTACGACAGTGGTAGCC 3000
TGCTTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAATCACTTAGTCTCTCTTTCGCGGATTTTITGACCTCATGCTGTACCATCGG
----- insert pLM1 -----
----- ORF pLM1 -----
T D S E K R S L A E S G L S V F S E S E E K A P K K L E Y D S G S
TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCATCCAAGGGTGGAGAATGAAAAAGCCCATCAGCCTGGG 3100
ACTTCTACCTTGGACCTGAAGATTACCGCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCACCTCTTGACTTTTTCGGGTAGTCGGACCC
----- insert pLM1 -----
----- ORF pLM1 -----
L K M E P G T S K V R R E R P E S C D D S S K G G E L K K P I S L G
CCACCTGGTTCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCATCACTCACACAGCCAGAGTGCCCTCAAAGTCGAGGCAACCT 3200
GGTGGGACCAAGGGACTTCTTCCCGTCTGGGGTGGACACCGACATTGAAGGGGTAGTGAGTGTGTCGGGTCTACGGGAGTTTCAGCGTCCGTTTGA
----- insert pLM1 -----
----- ORF pLM1 -----
H P G S L K K G K T P P V A V T S P I T H T A Q S A L K V A G K P
GAGGGCAAGCTACAGACAAGGTAAGCTTGCAAGTGAAGAATACTGGGCTCAACGCTCCTCTCTGATGCTGGTCGGGACCGCTGAGTGATGCTAAGA 3300
CTCCCGTTTCGATGTC TGTTCCTTCGAACTGCAAGTCACTTATGACCCGAGGTTCGAGGAGGAGACTACGACAGCCCTGGCGGACTCACTACGATCTT
----- insert pLM1 -----
----- ORF pLM1 -----
E G K A T D K G K L A V K N T G L O R S S S D A G R D R L S D A K

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Page 4

AGCCCCC TC GG G C A T T G C T C G C C C T C C A C T T C G G G A T C C T T C G G C T A C A A G A G C C T C C T C C T G C C A C A G G C A C A G C C A C T G T C A T G C A A A C T G G T G G
TC GG G G G A G C C C G T A A C G A G C G G G A G G T G A A G C C C T A G G A A G C C G A T G T T C T T C G G A G G A G G A C G G T G C C G T G C G G T G A C A G T A C G T T T G A C C A C C 3400

insert pLM1

ORF pLM1

K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G G

T T C A G C C A C T C T C A G A A G A T C C A G A G T C C T C A G G C A T C C C T G T C A A G C C A G T A A A T G G G C G C A A G A C T A G C T T A G A T G T T T C C A A C A G C G C A G A G C C A
A A G T C G G T G A G A G T C G T T C T A G G T C T T C A G G A G T C C G T A G G G A C A G T T C G G T C A T T T A C C C G C G T T C G A T C G A A T C T A C A A G G T T G T C G C G T C T C G G T 3500

insert pLM1

ORF pLM1

S A T L S K I O K S S G I P V K P V N G R K T S L O V S N S A E P

G A T T C C T G G C T C C T G G A G C C G T T C T A A C A T C C A G T A C C G A G C C T G C C C G G C C A G C C A A G T C A A G T T C T A T G A G C G T G A C C G G C G G G G G G T G G A C
C C T A A G G A C C G A G G A C C T C G G G C A A G A T T G T A G G T C A T G G C T C G G A C G G G G C C G G T C G G T T C A G T T C A A G A T A C T C G C A C T G G C C G C C C G C C C A C C T G 3600

insert pLM1

ORF pLM1

G F L A P G A R S N I O Y R S L P R P A K S S S H S V T G G R G G

C T C G C C C T G T G A G C A G C A G A T T G A C C C A G T C C C T C A G C A C C A A G C A G G A G G C C T T A C G C C T T C C A G A C T G A A G G A G C C T A C C A A G G T A G C C A G T G G
G A G C G G A C A C T C G T C G T C G T A A C T G G G G T C A G A G G A G T C G T G G T T C G T C C C T C C G G A A T C G G A A G G T C T G A C T T C C T C G G A T G G T T C C A T C G G T C A C C 3700

insert pLM1

ORF pLM1

P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S G

G C G G A C C A C T C C A G C C C T G T C A A T C A G A C A G A T C G G G A A A G G A G A G G C C A A G C C A A G G C A G T G G C C T T G G A C T C A G A C A A C A T C T C C T T G A A G A G T
C C C C T G G T G A G G T C G G G A C A G T T A G T C T G T C T A G C C C T T T C C T C T T C C G G T T C G G T T C C G T C A C C G G A A C C T G A G T C T G T T G T A G A G G A C T T C T C A 3800

insert pLM1

ORF pLM1

R T T P A P V N O T D R E K E K A K A K A V A L D S D N I S L K S

A T T G C T C C C A G A G A G T A C C C A A G A C C A A G C A A G C C A C C C C A C A G C C A C C A A G C T G G C A G A G T G C C A C C A A C C C C T C T C A G G G C C A C A G C G A A G A
T A A C C G A G G G G T C T C T A T A G G G T T C T T G G T C G T T C G G T G G G G T G T C G G T G G T T C G A C C G T C T C G A C G T G G T T G G G G A G A T C C C G G T G T C G C T T C T 3900

insert pLM1

ORF pLM1

I G S P E S T P K N O A S H P T A T K L A E L P P T P L R A T A K

G T T T G T C A A C C A C C C T C A C T A G C C A A T C T T G A C A A G G T A A C T C C A C A G T C T G G A T C T A C C A T C A C C A G T A T A C C A C C C A T G C T T C A A G G T C C C
C A A A C A G T T T G T G G G A G T A T C G G T T A G A A C T G T T C C A G T T G A G G T T G T C A G A C C T A G A T G G T A G T A G G T C A T A T G T G G G T A C G A A G T T T C C A G G 4000

insert pLM1

ORF pLM1

S F V K P P S L A N L D K V N S N S L D L P S S S D T T H A S K V P

A G A T C T G C A T G C T A C A A G C T A G C A T C T G G G G C C C T C C C C T C C T G C T T C A C C C C A G T C C G G C A C C C A C C T C C A A T A T T A A C T A G C C A G C T T C C C
T C T A G A C G T A C A T G T T C G A G T C G T A G A C C C C C G G A G A G G A A G G A C A A G T G G G G T C A G C C G T G G G T A G G A G T T A A T T G A G T C G G T C A A G A G G 4100

insert pLM1

ORF pLM1

D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F S

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 5

CAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGATGTACCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGC
GTCCCGGACCTCGATTACTACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTCTCTCTACCTCAGGGAGGTCACG 4200

insert pLM1

ORF pLM1

U G L E L M S G F S V P K E T R M Y P K L S G L H R S H E S L O M

CAATGAGCCTCCCAAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTGCTGCTCCACAGAAAGAGACGGAAGAGCTGACTTG
GTTACTCGGAGGGGTCACGAAGGGGTCGTATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTCTGCTCTCTGACTGAAC 4300

insert pLM1

ORF pLM1

P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T V

GAGTGAAGCCCAAGAGCTGGGCACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAG
CTCACCTTCGGGGTCTGACCCGTTGACCTGTCTATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTCCCGAGTCCATGGTCGAAGTCAGGGTCCCTCTC 4400

insert pLM1

ORF pLM1

S G S P R A G O L D S N Q R O R N T L P K K G L R Y O L O S Q E E

ACCAAGGAGAGGGACATTCACATACCATTTGGTGGGCTGGCTGAATCCGATGACCACTCAGAGCTGCCTTCTCCCCCTGCATTCCTCATGCTCTGAGTG
TGGTTCCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGCTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCAC 4500

insert pLM1

ORF pLM1

T K E R R H S H T I G G L P E S D D O S E L P S P P A L P M S L S

CAAAGGGCCAACCTTACCAACATAGTGAGTCCCACTGCGGCCACCAAGCAAGCAATCACCCGCTCCAACAGCATCCCCACCCAGGCGGCTTCGAGCT
GTTTCCCGGTGAATGGTTGTATCCTCAGGGTGACCGGGTGGTGGGTTCTTAGTGGGCGAGGTTGTCTGAGGGGTGGGTGCTCCGCGGAAGCTCGA 4600

insert pLM1

ORF pLM1

A K G O L T N I V S P T A A T T P R I T R S N S I P T H E A A F E L

GTACAGCGGCTCCCAATGGGAGCACCTGTCTCTGGCCGAGAGACCAAGGGAATGATTGGTCAAGTACCTTCCGAGACCCACGGACGATGTTTAC
CATGTCCGCGAGGGTTTACCCCTCGTGGGACAGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCTTAGGAAGGCTCTGGGGTGCCTGCTACAAGTG 4700

insert pLM1

ORF pLM1

Y S G S O M G S T L S L A E R P K G H I R S G S F R D P T D D V H

GGCTCAGTGTCTCCCTGGCTCCAGTGCTTCTCCACCTACTCTCAGCTGAGGAGAGGATGCAATCTGAGCAAAATCCGGAAGCTTCGTAGGGAAGTGG
CCGAGTCAAGACAGGACCGAGGTCACGGAGGAGTGGATGAGGAGTCCAGTCTCTCTACGTAGTACGTTTGGGCTTCGAGCATCCCTTGACC 4800

insert pLM1

ORF pLM1

G S V L S L A S S A S S T Y S S A E E R H O S E O I R K L R R E L

AATCATCCAGGAAAAAGTGCCACCTTGACGTCTCAGCTTCTGCCAATGCTAATCTGGTGGCTGCTTTGAGCAGAGCTGGTGAATATGACATCCCG
TTAGTAGGGTCTTTTTCACCGGTGAACTCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAAATCTGCTCGGACCACTTATATGTAAGGGC 4900

insert pLM1

ORF pLM1

E S S O E K V A T L T S O L S A N A N L V A A F E O S L Y N M T S R

Tuesday, 18 November 1997 13:57
lig 54 pLM1 (1 > 8285) Site and Sequence

Page 6

```
CCTGCGACACCTGGCAGAGACGGCCGAGGAGAAAGGACACTGAGCTGCTGGATTGCGAGAAACCATAGACTTCTGAAGAAAAAGAACCTGAGGCCAG
GGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCTGTGACTCGACGACCTAAACGCTCTTGGTATCTGAAAGACTTCTTTCTTGAGACTCCGGGTC 5000
-----insert pLM1-----
-----ORF pLM1-----
L R H L A E T A E E K D T E L L O L R E T I D F L K K K N S E A Q
GCAGTCATTCAGGAGGCCCTTAATGCCTCAGAAACCAACCCAAAGAACTTCGGATCAAGAGACAAAACCTCAGATAGCATCTCAAGCCTCAACAGCA
CGTCAGTAAGTCCCTCGGAATTACGGAGTCTTGGGTGGGTTCTTGAAGCCTAGTTCTCTGTTTGGAGGATCTATCGTAGAGTTCGGAGTTGTCGT 5100
-----insert pLM1-----
-----ORF pLM1-----
A V I O G A L N A S E T T P K E L R I K R O N S S D S I S S L N S
TCAGTAGCCATTCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGC
AGTGATCGGTAAGGTCGTAGCCGTCGTCCTTACGACTACGCTTTTCTCTTTTCTTCTCAACCCAGATACCGAAGCTTCAAGGAAGTTGTTTCG 5200
-----insert pLM1-----
-----ORF pLM1-----
I T S H S S I G S S K D A D A K K K K K K S V V Y E L R S S F N K A
GTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAAACTACAG
CAAGTCATATTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTGATGTC 5300
-----insert pLM1-----
-----ORF pLM1-----
F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L Q
CATGTTCCACAGAGACTGCTTACCCCTCCATCAAGTCTCCACCTTGTCTCCGTTGGGCACATGATGTACCCGAGGGCCCTGCTCACCAGCCCCCACA
GTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTTCAGGAGGTGGAACAGGAGGACCCGTGACTACAGTGGCTCCCGGACGAGTGGGTGCGGGGGTGT 5400
-----insert pLM1-----
-----ORF pLM1-----
H G S T E T A S P S I K S S T L S S V G T D V T E G P A H P A P H
CTAGGCTGTTCCATGCAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAATGAAGCTTACAGACAT
GATCCGACAAGGTACGTTTACTCTCTCTCTCGGTCCTCTCTTCTTCCATAGCCTCGACGCGAGACTCGATACCTCTCTCTTACTTCGAAIGTCTGTA 5500
-----insert pLM1-----
-----ORF pLM1-----
T R L F H A N E E E E P E K K E V S E L R S E L V E K E H K L T D I
CCGC TTGGAGGCCCTCAACTCTGCCACCACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACC TGCTGAAAGCAGAGAATGAC
GGCGAACC TCCGGGAGTTGAGACGGGTGGT TGACCTAGTCGAASCCCTCTGGTACGTGTGTACGTCAACC TCCACCTGGACGACTTTCGTCCTTACTG 5600
-----insert pLM1-----
-----ORF pLM1-----
R L E A L N S A H O L D O L R E T H H N H O L E V D L L K A E N D
CGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGATTATCTTCCCCACGCCGCTCC TAGGCTTGGCAC
GCTGACTTCCATCGGGGTCCGGGAGTAGTCCGAGGTGAGGTCCCTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTG 5700
-----insert pLM1-----
-----ORF pLM1-----
R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S L G L A
```

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

TCACCCATTCTTCGGCCCCAGCTTGCAGACACAGACC TGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAGGAGGAAGTGACCC TCGGGTGGT
AGTGGGTAAGGAAGCCGGGTGACAACGTC TGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCTCCTTCACTGGGAGGCCACCA 5800

insert pLM1

ORF pLM1

L T H S F G P S L A D T D L S P H D G I S T C G P K E E V T L R V V
GGTGAGGATGCCCCGAGCAGCATCATCAAAGGGGACTTGAAGCAGCAGGAATTC TCTCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATG
CCAC TCCTACGGGGGCGTCGTAGTAGTTTCCCTGAACTTCGTCTCTTAAGAAGGACCCGACATCGTTCAGTCACCTTTCAACTGACCTTCTAC 5900

insert pLM1

ORF pLM1

V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D V K H
CTGGATGAAGCTGTTTCCAAGTGTCAAGGACTATATTCTAAAATGGACCCAGCC TC TACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCA
GACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAGATTYTACCTGGGTCGGAGATGGGACCCCTGATTCGTGACTCAGGTAGGTACCGATGTCGT 6000

insert pLM1

ORF pLM1

L D E A V F Q V F K D Y I S K H D P A S T L G L S T E S I H G Y S
TCAUCCACGTGAACAGAGTGTGGATGCAGAGCCCCGAGATGCCTCTTGGCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGA
AGTCTGGTGCACTTTGTCTACAACCTACGTC TCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCT 6100

insert pLM1

ORF pLM1

I S H V K R V L D A E P P E H P P C R R G V N N I S V S L K G L K E
GAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGTCTCTCG
CTTACGCACTGTCGGACCAAGCTCTCGGACTAGGGGTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGC 6200

insert pLM1

ORF pLM1

K C V D S L V F E T L I P K P N M Q H Y I S L L L K H R R L V L S
GGCCCGACGGGACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCT
CGGGGTCGCGTGGCTTCTTGGATGGAC TGGTTAGGAACCGGCTCATGGACCACTCGGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGA 6300

insert pLM1

ORF pLM1

G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T
TCAACATGCACAGCACTCTTGAAGGATCTGCAACTGTATCTTCCAACCTAGCCAACAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTGGT
AGTTGTACGTGGTCTGCAAGCTTCTAGACGTGACATAGAAAGGTGGATCGGTTGGTCTATCTGGCCCTTTGCTCTTAACCCCTACACGGGGACCA 6400

insert pLM1

ORF pLM1

F N M H D Q S C K D L Q L Y L S N L A N Q I D R E T G I G D V P L V
GATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTGGTCAATGGGGCCCTACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC
CTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCGGGAGTGGACGTTCATAGTATTACAGGGATATAATATCCATGG 6500

insert pLM1

ORF pLM1

I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G T

Page 5

BNSDOCID: <WO__9824810A2 | >

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

AGSAGGGACAGGTTCTTGGTGTGTACCTTTGAGAACTTCTTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGC 7400
TCCTCCCTGTCCAAGAACCACGACATGGAACCTCTTGAAGGATCCTTCTTACCACCCACCGCAAACCTTGAACACGGGGGATTTGTGTAATGACCG
-----insert pLM1-----
G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G
CTCTCTAATGACTTTGGGAAAAAGATGATCTGGGTCTTCCCTTGACTTCTTGTTCATTACAACTCCTGGGCTTCTGGGGAGGGGTTTCAGAAAA 7500
GAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGAACTGAAGAACAAGTTAATGTTTGGAGACCCGAAAGACCCCTCCCAAGTCTTT
-----insert pLM1-----
L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K
CATCAAAACACTGCAGCAGTTCCCCGAATTCAGCTTGGACTTAACAGGCTGAACCTTGCTCAAAGAAGCCGAATTCAGCACACTGGCGGGCGGTACT 7600
GTAGTTTTGTGACGTCGTAAGGGGCTTAAGTCGAACCTGAATTTGGTCCGACTTGAACGAGTTTCTTCGGCTTAAGGTCGTGTGACCGCGGCAATGA
-----insert pLM1-----
T S K H C S S S P E F S L D L T R L N L L K R S R I P A H V R P L L
AGTTCTAGAGCGGCGCCACCGGGTGGAGCTCAATTGCCCCATAGTGAGTCGTATTACGCGCGCTCACTGGCGGTGTTTTACAACGTCGTGACTGG 7700
TCAAGATCTGCGCGCGGTGGCGCCACTCGAGGTTAAGCGGGATATCACTCAGCATAATGCGCGGAGTGACCGGCAGCAAAATGTTGAGCACTGACC
→
V L E R P P P R V S S N S P Y S E S Y Y A R S L A V V L O R R D V
GAAACCCCTGGGCTTACCAACTTAATCGCCTTGACGACATCCCCCTTCCGCGAGTGGCGTAATAGCGAAGAGGCCCGACCGATCGCCCTTCCCAAC 7800
CTTTTGGGACCGCAATGGGTGAATTAGCGGAACGTCGTGTAGGGGGAAGCGGTGACCGCATTATCGCTTCTCCGGCGTGGCTAGCGGGAAGGGTTG
E N P G V T Q L N R L A A H P P F A S V R N S E E A R T D R P S Q
AGTTGCGCAGCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGTGTGGTGTACGCGCAGCTGACCGCTACACTTGGCAG 7900
TCAACGCGTCGGAATTACCGTTACCTGCGGGGACATCGCCGCGTAATTCGCGCGCCACACCAATGCGCGTGGCACTGGCGATGTGAACGGTC
Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S
CGCCCTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCTTCGCCACGTTCCGCGGCTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTC 8000
GCGGGATCGCGGGCAGGAAAGCGAAAGAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGAGTTGAGATTAGCCCCGAGGGAATCCCAAG
A L A P A P F A F F P S F L A T F A G F P R Q A L N R G L P L G F
CGATTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTGA 8100
GCTAAATCAGGAAATGCCGTGGAGCTGGGGTTTTTGAACATAACCCAC TACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAC T
R F S A L R H L D P K K L D . G D G S R S G P S P . . T V F R P L
CGTTGGAGTCCAGTTCCTTAATAGTGGACTCTTGTTCCAAAC TGAACAACACTCAACCTTATCTCGGTCTATCTTTTGATTATAAGGGATTTTGCC 8200
GCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAGGTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAAC TAAATATCCCTAAACCGG
T L E S T F F N S G L L F O T G T T L N P I S V Y S F D L . G I L P
GATTTCGGCCTATTGGTTAAAAAATGAGCTGATTAAACAAAAATTTAACGCAATTTTAAACAAATATTAAACGCTTACAATTTAG 8285
CTAAAGCCGGATAACCAATTTTACTCGACTAAATTTGTTTAAATTCGCTTAAATTTGTTTATAATTGCGAATGTTAAATC
I S A Y V L K N E L I . O K F N A N F N K I L T L T I .

Tuesday, 18 November 1997 13:57

fig 55 pCB251 (1 > 8197) Site and Sequence

Enzymes: All 148 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

GACGGATCGGGAGATCTCCCGATCCCCATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT;
CTGCC TAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCAATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V
GGAGGTCGCTGAGTAGTGC GCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTCGS
CCTCCAGCGACTCATCAGCGCTCGTTTAAATTCGATGTTGTTCCGTTCCGAACGGCTGTAAACGTACTCTTAGACGAATCCCAATCCGCAAAACGC
G G R . V V R E Q N L S Y N K A R L D R O L H E E S A . G . A F C
CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCGGTCTATATGCGCAACTGTAAC TAATACTGATCAATAATTATCATTAGTTAATGCCCGAGTAATCAAGTATCGGGTATAT
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y
TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCGCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGCGGGTGTCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A W L T A Q R P P P I D V N N D V C S H S
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
TTGCGGTATCCCTGAAAGGTAACGCGATTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCCGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A
CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTCTCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCACTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCAATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT
P Y . R O . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H
TCGCTATTACCATGGTGATCGGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTTGACTCAGGGGATTTCAGGTCTCCACCCATTGACGTCAA
AGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCGCACCTATCGCCAAAC TGAGTGGCCCTAAAGGTTGAGAGGTGGGGTAACGCGATT
R Y Y H G D A V L A V H O W A V I A V . L T G I S K S P P H . R Q
TGGGAGTTTGTGTTGGCACAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAG
ACCTCAAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTGTTGAGGCGGGTAACGCGTTTACCGCCATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G
GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCACCTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCAAGCTGGCTAGC
CAGATATATTCGCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGACCGCATCG
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A
GTTTAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG
F K L K L T M G G S H H H H H G M A S M T G G Q Q M G R D L Y D
GATGACGATAAGGTACCGGATCTTCCGAGACCCACGAGCATGTTACGGCTCAGTGCTGTCCCTGGCTCCAGTGCCTCCTCCACCTACTCCTCAG
CTACTGCTATTCCATGGGCTAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTCACGGAGGAGTGGGATGAGGAGT
C D D K V P G S F R D P T D D V H G S V L S L A S S A S S T Y S S
pCB251 insert = U2

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 2

CTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCTTCGTAGGGAAC TGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCACGTTTCTGCCAA
GACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTAGTAGGGTCCTTTTCACCGGTGGAAC TGCAGAGTCGAAAGACGGT 120

pCB251 insert = U2

U2 ORF

A E E R M Q S E Q I R K L R R E L E S S Q E K V A T L T S Q L S A N

TGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTG
ACGATTAGACCACCGACGAAAACTCGTCTCGGACCACCTTATACTGTAGGGCGGACGC TGTGGACCGTCTCTGCCGGCTCCTCTCTCTGTGACTCGACGAC 130

pCB251 insert = U2

U2 ORF

A N L V A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L

GATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAAC TCTGAGGCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCAAAGAAC
CTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTCCTTG 140

pCB251 insert = U2

U2 ORF

D L R E T I D F L K K K N S E A Q A V I Q G A L N A S E T T P K E

TTCGGATCAAGAGACAAAACCTCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAA
AAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTT 150

pCB251 insert = U2

U2 ORF

L R I K R Q N S S O S I S S L N S I T S H S S I G S S K D A D A K I

GAAGAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATCTCGGATATA
CTCTTTTTTTCTCAACCCAGATCTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATAT 160

pCB251 insert = U2

U2 ORF

K K K K S W V Y E L R S S F N K A F S I K K G P K S A S S Y S D I

GAGGAGATTGCTACACCGACTCTTCAGCCCCCTCATCCCCAACTACAGCATGGTTCTACAGAGACTGCTTACCCTCCATCAAGTCCCTCCACCTTGT
CTCTCTAACGATGTGGGC TGAGAAGTCGGGGGAGTAGGGGTTTGATGTCGTACCAAGATGTC TCTGACGAAGTGGGAGGTAGTTTACGGAGGTGGAACA 170

pCB251 insert = U2

U2 ORF

E E I A T P D S S A P S S P K L Q H G S T E T A S P S I K S S T L

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 3

CCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTACCCAGCCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAGAAAGAGGAGGT
GGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTGGTCTCTTCTCTCTCCA 180

pCB251 insert = U2

U2 ORF

S S V G T D V T E G P A H P A P H T R L F H A N E E E E P E K K E V

ATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGAGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAG
TAGCCTCGACGCAGACTCGATACCCCTTCTCTTACTTCAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTC 190

pCB251 insert = U2

U2 ORF

S E L R S E L V E K E M K L T D I R L E A L N S A H Q L D Q L R E

ACCATGCAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGG
TGGTACGTGTTGTACGTCAACCTCCACC TGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCC 200

pCB251 insert = U2

U2 ORF

T M H N M Q L E V D L L K A E N D R L K V A P G P S S G S T P G Q

TCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTCACCATTCCTTCGGCCCCAGTCTTGCAGACACAGACCTGTACCCAT
AGGGACCTAGTAGACGTAATAGAAGGGTGGCGGAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCCGGGGTCAGAACGCTGTGTCTGGACAGTGGGT 210

pCB251 insert = U2

U2 ORF

V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D L S P H

GGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTACCCTCCGGGTGGTGGTGGATGCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAG
CCTACCGTAGTCATGAACACCAGGTTTCTCTTCTACTGGGAGGCCACCACTCTCTACGGGGGCGTCGTGTAGTAGTTTCCCTGAACCTTCGTCTCT 220

pCB251 insert = U2

U2 ORF

D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q

GAATTCTTCTGGGCTGTAGCAAGGTACGTGGAAGGTTGACTGGAAGATGCTGGATGAAGCTGTTTCCAAGTGTTCAGGACTATATTTCTAAATGG
CTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTTCAAGTTCTGATATAAAGATTTTACC 230

pCB251 insert = U2

U2 ORF

E F F L G C S K V S G K V D W K M L D E A V F Q V F K D Y I S K H

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 4

ACCCAGCCTCTACCCCTGGGACTAAGCAC TGAGTCCATCCATGGCTACAGCATCAGCCACGTGAACGAGTGTGGATGCAGAGCCCCGAGATGCCTCC
TGGGTCGGAGATGGGACCCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGACAC TTGCTCACAACTACGTCTCGGGGGGCTCTACGGAGG 240

pCB251 insert = U2

U2 ORF

D P A S T L G L S T E S I H G Y S I S H V K R V L D A E P P E M P F

TTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGAGCTGATCCCCAAGCCGATG
AACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCAGACTTCCTCTTTACGCAGCTGTCTGGACCACAAGCTCTGCGACTAGGGGTTCGGCTAC 260

pCB251 insert = U2

U2 ORF

C R R G V N N I S V S L K G L K E K C V D S L V F E T L I P K P M

ATGCAGCACTACATAAGCCTCCTGCTGAAGCACC GGCGCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGT
TACGTCGTGATGATTTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTCTGCCGTTCTGGATGGACTGGTTAGCGAACC GGCTCA 280

pCB251 insert = U2

U2 ORF

M Q H Y I S L L L K H R R L V L S G P S G T G K T Y L T N R L A E

ACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAA
TGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTGAGAAGCTTCTTAGACGTTGACATAGAAAGGTT 290

pCB251 insert = U2

U2 ORF

Y L V E R S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N

CCTAGCCAACCAGATAGACCGGAAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTC
GGATCGGTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGACCCTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACCTCAACCAAG 300

pCB251 insert = U2

U2 ORF

L A N O I D R E T G I G D V P L V I L L D D L S E A G S I S E L V

AATGGGGCCCTCACCTGCAAGTATCATAAATGTCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCT
TTACCCCGGGAGTGGACGTTTATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTGGTACCGAACGTGAACTCGA 310

pCB251 insert = U2

U2 ORF

N G A L T C K Y H K C P Y I I G T T N O P V K M T P N H G L H L S

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 5

TCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCC TGGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGC
AGTCC TACAAC TGAAGAGGTGTTGCACCTCGGTCGGT TACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGAGTTACG 300

pCB251 insert = U2

U2 ORF

F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A

CAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCTTGAGAAGCACAGCACCTCAGACTTCTCATC
GTTGTTCTTCTCGACGAAGCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAG 310

pCB251 insert = U2

U2 ORF

N K E E L L R V L D V V P K L V Y H L H T F L E K H S T S D F L I

GGCCCTTGCTTCTTCTGTCGTGTCCTTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACTCTATCATTCCTATCTACAGGAAG
CCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCC TGGACCAAGTAAC TGGACACCTTGTTGAGATAGTAAGGGATAGATGTCCTTC 320

pCB251 insert = U2

U2 ORF

G P C F F L S C P I G I E D F R T V F I D L V N N S I I P Y L Q E

GAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGA
CTCGGTTCTACCTATTTCCAGGTACCTGTCTTTCGACGAACCCTCTGGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCT 330

pCB251 insert = U2

U2 ORF

G A K D G I K V H G Q K A A V E D P V E V V R D T L P W P S A Q Q D

CCAATCAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCT
GGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGA 340

pCB251 insert = U2

U2 ORF

Q S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S

CTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCTGGACCCCAACCTTC
GACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAACCTCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAG 350

pCB251 insert = U2

U2 ORF

L D S D P L M A M L L K L Q E A A N Y I E S P D R E T I L D P N L

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1>8197) Site and Sequence

Page 6

AGGCAACACTTTAAGGGTTCCGCAATCACTGTCAACCCCGGACAGCAGAAGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCTCTCCTCTTTTCAGAG 360X
TCCGTTGTGAAATTCCCAAGCCGTTAGTGACAGTGGGGGCCGTGTCGCTTTCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTC

pCB251 insert = U2

U2 ORF

Q A T L . G F G N H C H P R T A E R V H Q L S . L L L S P L L F Q S

CACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGAGATGAAAGAGGAGGGACAGGTTCCTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGG 370X
GTGACCGAGAGGTCGGGGTCTCCTCTTGTCTCCCTCTCTCTACTTCTCCTCCCTGTCCAAGAACCACGACATGGAACCTCTTGAAGGATCCTTCC

pCB251 insert = U2

T G S P A P G G E Q E G G G D E R G G T G S V C C T F E N F L G R

AATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAAACACATTTACTGGCCTCTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCT 380X
TTACCACCCACCGCAAACCTTGAACACGGGGGATTGTGTAATGACCGGAGGAGATCTCCCGGGCAAATTGGGCGACTAGTCGGAGCTGACACGGA

pCB251 insert = U2

N G G V A F G N L C P L N T F T G L L . R A R L N P L I S L D C A

TCTAGTTGCCAGCCATCTGTTGTTTCCCCCTCCCGTGCCTTCCTTGACCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGAGGAAATG 390X
AGATCAACGGTCGGTAGACAACAACGGGGAGGGGACGGAAGGAAC TGGGACCTTCACGGTGAGGGTGACAGGAAAGGATTATTTACTCCTTTAAC

F . L P A I C C L P L P R A F L D P G R C H S H C P F L I K . G N C

CATCGCATTTGCTGAGTAGGTGTCTATTCTATTCGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGG 400X
GTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCACCCACCCCGTCTGTGCTTCCCTCTCTAACCTTCTGTTATGCTGCGTACGACCCCT

I A L S E . V S F Y S G G V G G A G Q Q G G G L G R Q . Q A C W G

TGGGTGGGCTCTATGGCTTCGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAGCGCGCGGGGTGTS 410X
ACGCCACCCGAGATACGAAGACTCCGCTTTCTTGGTCGACCCCGAGATCCCCATAGGGGTGCGCGGGACATCGCCGCGTAATTCGGCGCGCCACAC

C G G L Y G F . G G K N Q L G L . G V S P R A L . R R I K R G G C

GTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCTTCTCTCGCCACGTTGCGCGGCTTTCCCG 420X
CACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCGGAAAGGGG

G G Y A Q R D R Y T C Q R P S A R S F R F L P F L S R H V R R L S F

GTCAAGCTCTAAATCGGGGATCCCTTTAGGGTTCGATTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTACGCTAGTGG 430X
CAGTTCGAGATTTAGCCCCGTAGGGAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTTGGGGTTTTTGAACATAATCCCACTACCAAGTGATCACC

S S S K S G H P F R V P I . C F T A P R P Q K T . L G . W F T . W

GCCATCGCCCTGATAGACGGTTTTTCCCTTTGACGTTGGAGTCCACGTCTTTAATAGTGGACTCTTGTTCCAAACGGAACAACACTCAACCCATC 440X
CGGTAGCGGGACTATCTGCCAAAAAGCGGAAACTGCAACCTCAGGTGCAAGAAATATACCTGAGAACAGGTTTGACCTTGTGTGASTTGGGATAG

A I A L I D G F S P F D V G V H V L . . W T L V P N W N N T Q P Y

TGGGTCTATTCTTTGATTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCGTG 450X
AGCCAGATAAGAAAACTAAATATTCCTTAAACCCCTAAAGCCGATAACCAATTTTTACTCGACTAAATTGTTTTAAATTCGCGTTAATTAAGACAC

L G L F F . F I R D F G D F G L L V K K . A D L T K I . R E L I L W

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 7

GAATGTGTGTCAGTTAGGGTGTGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAG 260
CTTACACACAGTCAATCCCACACCTTTACAGGGTCCGAGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTT
N V C Q L G C G K S P G S P G R Q K Y A K H A S Q L V S N O V W I
GTCGCCAGGCTCCCCAGCAGGCAGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCGCCCTAACTCCGCCCATCCGCCCTAACT 270
CAGGGGTCCGAGGGTCTGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGA
V P R L P S R Q K Y A K H A S Q L V S N H S P A P N S A H P A P N
CCGCCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGT 280
GGCGGGTCAAGGCGGGTAAAGAGGCGGGTACCGACTGATTAATAAATAAATACGTCCTCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCA
S A Q F R P F S A P W L T N F F Y L C R G R G R L C L . A I P E V V
GAGGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCG 290
CTCTCCGAAAAAACCTCCGGATCCGAAAAACGTTTTTCGAGGGGCCCGAACATATAGGTAAGGCTTAGACTAGTTCTCTGCTCTACTCTAGCAAAAGC
R R L F W R P R L L Q K A P G S L Y I H F R I . S R D R M R I V S
CATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGAT 300
GTACTAAGTTGTTTACCTAACGTGCGTCCAGAGGCCGCGCAACCCACCTCTCCGATAAGCCGATACTGACCCGTTGTTGTTGTTAGCCGACGAGACTA
H D . T R V I A R R F S G R L G G E A I R L . L G T T D N R L L .
GCCGCCGTGTTCCGGCTGTGACGCGAGGGGCCCGGTTCTTTTGTCAAGACCGACCTGTCGGTGCCTGAATGAAGTGCAGGACGAGGCAGCGCGGC 310
CGGGGCGACAAGGCGAGCTCGCGTCCCGCGGGCCAAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCGCGCCG
C R R V P A V S A G A P G S F C Q D R P V R C P E . T A G R G S A A
TATCGTGGCTGGCCACGACGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCATGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTCCCGGGCA 320
ATAGCACCGACCGGTGCTGCGCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACGACGATAACCCGCTTCACGGCCCCGT
I V A G H D G R S L R S C A R R C H . S G K G L A A I G R S A G A
GGATCTCTGTCTCATCTACCTTGCTCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC 330
CCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTTCATAGGTAGTACCGACTACGTTACGCCGCCGACGATGCGAAGTAGGCGGATGGAGGGTAAG
G S P V I S P C S C R E S I H H G . C N A A A A Y A . S G Y L P I
GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC 340
CTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGACCTACCTTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTCGTAGTCCCGAGCGCG
R P P S E T S H R A S T Y S D G S R S C R S G . S G R R A S G A R A
CAGCCGAACGTGTTGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCGTGACCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA 350
GTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTT
S R T V R Q A Q G A H A R R R G S R R D P W R C L L A E Y H G G I
TGGCCCTTTTCTGGATTATCAGCTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGC 360
ACCGGCGAAAAGACCTAAGTAGCTGACACCGGCGGACCCACACCGCTGGCGATAGTCTGTATGCAACCGATGGGCACTATAACGACTTCTCGAACCG
W P L F W I H R L W P A G C G G P L S G H S V G Y P . Y C . R A W
GGCGAATGGCTGACCGCTTCTCGTGTTCAGGATGCGCGCTCCGATTCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGG 370
CCGCTTACCGGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAATGCTCAAGAAGACTCGCC
R R M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 8

GACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCACGAGATTCGATTCCACCGCCGCTTCATGAAAGGTTGGGCTTCGGAATC
CTGAGACCCCAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAAACCCGAAGCCTTAG 520
T L G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I
GTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCGCCCACCCCAACTGTTTATTGCAGCTTATAATGGTTACA
CAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAATATTACCAATGT 530
V F R D A G W M I L Q R G D L M L E F F A H P N L F I A A Y N G Y
AATAAGCAATAGCATCACAAATTTCAAAATAAGCATTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTG
TTATTCGTTATCGTAGTGTAAAGTGTATTTTCGTAAGGAGTACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATAGTACAGAC 540
K . S N S I T N F T N K A F F S L H S S C G L S K L I N V S Y H V C
TATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATACGAGCC
ATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTGTATGCTCGG 550
I P S T S S . S L A . S W S . L F P V . N C Y P L T I P H N I R A
GGAAGCATAAAGTGTAAGCCTGGGGTGCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
CCTTCGTATTTACATTTTCGGACCCACGGATTACTCACTCGATTGAGTGTAATTAAACGCAACGCGAGTGACGGGCGAAAGGTACGCCCTTTGGACAGCA 560
G S I K C K A V G A . . V S . L T L I A L R S L P A F Q S G N L S
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGGCTATTGGGCGCTCTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGT
CGGTGCGAGTAATTACTTAGCCGTTTGGCGGCCCTTCGCCCAACGCATAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCA 570
C Q L H . . I G O R A G R G G L R I G R S S A S S L T D S L R S V V
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAAGGAAGAACATGTGAGCAAAAGGCCAG
AGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCGCCATTATGCCAATAGGTGTCTTAGTCCCTATTGCGTCTTCTTGTACACTCGTTTCCGGTC 580
R L R R A V S A H S K A V I R L S T E S G D N A G K N M . A K G Q
CAAAAGGCCAGGAACCGTAAAGGCGCGGTTGCTGGCGTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGT
GTTTTCCGGTCTTTGGCATTTTCCGGCGCAACGACCGCAAAAGGTATCCGAGGCGGGGGAGTGCCTGAGTGTGTTTTAGCTGCGAGTTCACTGCCA 590
O K A R N R K K A A L L A F F H R L R P P D E H H K N R R S S Q R
GGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGT
CCGCTTTGGGCTGTCCTGATATTTCTATGGTCCGCAAGGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCTATGGACAG 600
V R N P T G L . R Y Q A F P P G S S L V R S P V P T L P L T G Y L S
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCAC
GCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTTCGACCCGACACAGTG 610
A F L P S G S V A L S Q C S R C R Y L S S V . V V R S K L G C V H
GAACCCCGGTTTCAGCCCGACCGCTGCGCCTTATCCSGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
CTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAAGTACAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCTGCTGGTGAC 620
E P P V Q P D R C A L S G N Y R L E S N P V R H D L S P L A A A T
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTATTGGTATCTG
CATGTCTAATCGTCTGCTCCATACATCCGCCACGATGCTCAAGAAGTTCACCAACCGGATTGATGCCGATGTGATCTTCTGTCTATAACCATAGAC 630
G N R I S R A R Y V G G A T E F L K V W P N Y G Y T R R T V F G I C

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 9

CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACCAACCACCGCTGGTAGCGGTGTTTTTTGTTTGCAGCAG 700
GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAAC TAGGCCGTTTGTGGTGGCGACCATCGCCACCAAAAAACAAACGTTTCGT
A L L K P V T F G K R V G S S . S G K Q T T A G S G G F F V C K Q
CAGATTACGCGCAGAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACTACGTTAAGGGATTTTGG 710
GTCTAATGCGCGTCTTTTTCTTAGAGTTCTCTAGGAACTAGAAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGTAGTGCAATTCCTTAAACCC
O I T R R K K G S Q E D P L I F S T G S D A Q W N E N S R . G I L
TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATAGTAACTTGGTCTGACAG 720
AGTACTCTAATAGTTTTCTTAGAAGTGGATCTAGGAAAATTTAATTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACAGACTGT
V M R L S K R I F T . I L L N . K . S F K S I . S I Y E . T W S D S
TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTCATCCATAGTTGCTGACTCCCCGTCGTGTAGATAACTACGATACGG 730
AATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCC
Y Q C L I S E A P I S A I C L F R S S I V A . L P V V . I T T I R
GAGGGCTTACCATCTGGCCCCAGTGCTGAATGATACCGCGAGACCCACGCTACCGGCTCCAGATTATCAGCAATAAACAGCCAGCCGGAAGGGCCG 740
CTCCCGAATGGTAGACCGGGTACGACGTTACTATGGCGCTCTGGGTGCGAGTGCGCGAGGCTCAAATAGTCGTTATTTGGTCGGTCGGCTTCCCGGC
E G L P S G P S A A M I P R D P R S P A P D L S A I N O P A G R A
AGCGCAGAAGTGGTCC TGCAACTTTATCCGCC TCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC GCA 750
TCGCGTCTTACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCAAGCGGTCAATTATCAACGCGTT
E R R S G P A T L S A S I Q S I N C C R E A R V S S S P V N S L R N
CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTTGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC 760
GCAACAACGGTAACGATGTCCGTAGCACACAGTGCAGCAGCAACCATACCGAAGTAAGTCGAGGCCAAGGGTGTCTAGTTCCGCTCAATGTACTAGG
V V A I A T G I V V S R S S F G M A S F S S G S Q R S R R V T . S
CCCATGTTGTGCAAAAAAGCGTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTATCACTCATGTTATGGCAGCACTGC 770
GGGTACAACACGTTTTTTCGCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACC
P M L C K K A V S S F G P P I V V R S K L A A V L S L M V M A A L
ATAATCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTG 780
TATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCAC TCATGAGTTGGTTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAAC
H N S L T V M P S V R C F S V T G E Y S T K S F . E . C M R R P S C
CTCTTGCCCGGCGTCAATACGGGATAATACCGGCCACATAGCAGAACTT TAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGS 790
GAGAACGGGCGCAGTTATGCCCTATTATGGCGCGGTGTATCGCTTGAAATTTTACAGAGTAGTAACCTTTTGCAAGAAGCCCGCTTTTGAGAGTTCC
S C P A S I R D N T A P H S R T L K V L I I G K R S S G R K L S R
ATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGCACCCCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGSGGTGAGCAAAAA 800
TAGAATGGCGACAACCTTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAGACCCACTCGTTTTT
I L P L L R S S S M . P T R A P N . S S A S F T F T S V S G . A K
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGG 810
GTCCTTCGTTTTACGGCGTTTTTCCCTTATCCCGCTGTGCTTTACAACCTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCC
T G R O N A A K K . G I R A T R K C . I L I L F L F Q Y Y . S I Y Q G

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 10

.TTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCACATTTCCCGAAAAGTGCCACCTGACGTC
AATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTATTTGTTTATCCCAAGGCGCGTGTAAGGGGCTTTTCACGGTGGACTGCAG 8197
Y C L M S G Y I F E C I . K N K Q I G V P R T F P R K V P P D V

Figure 56

Monday, 1 December 1997 14:13
fig 56 pNP10

Page 1

10 20 30 40 50 60 70
AACCTTGCATGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCEACC TCGAGGATCAGAAAGAAAT 70
TGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTTAATTTTGAGTTTACGAC TA 140
CAAAAATGTGTTCCTTAAAGAAC TAICTTCGAC TTAGAGCTATCTGTATGACTAGTTGTTGAGTGATT TT 210
TCAT TGAGAAAATATTAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCG 280
ATCAACTAGATCTTACAAAACCTTGAATACAATTCCATTTTCAGATTACCTTCGCCACGTGTCGCCACGT 350
360 370 380 390 400 410 420
CAGCAACCGCTTCAGCAACTAACCCAAATTCCTTCTCCACAAATGTCAACATCCAGGC TTAGACTCC 420
ACAGTCAAGAAATATCGAAAATTGGTAAGAAATTTATTTTGAGCTCAAACTTGTATAAAATGCCAGAAAA 490
GAAGAIGATAAAAAATGAGTTTTTTTGCAAAACCTCCACCTTTATGCTCTAATAAGACGGCTTATATCT 560
CAATTTTCTTGAGTTTTATCAAAAAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAAATTTTG 630
TCAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAAAGTGAAGGACCGCAAG 700
710 720 730 740 750 760 770
TCATACCTGGAATTAATGATCTGAAACAAATTTGTGCTATTCTCAAAATGTTTAAAGACATGTTTGAAGAT 770
TTTTTCAAATTCGCAC TATGTTTCAGAACCTTCCTTTTGTATGAAAAAGTAAAAAATAACTATTTCAAAC 840
CCTCACCGCCACCAATGTTTCAACTCTTAATTTTATAAAATTTTGCATTTACAATAAGCCCTCCCTTGC 910
CCGAAAAGTGGCCACCAAAAATCAATTTTTCGGCTTCATAATGACTTTTAAATTGATGTGAGAAAACACAG 980
AAGAGGCTAACTAAATTCACACGGGATAGGTGTTCCTCTTCTCCCTCCCTCTCCCTCTCCCTCTCCCT 1050
1060 1070 1080 1090 1100 1110 1120
TCCATCTCCAACAACAACAAATTTTCAATTTCTGTGCTCCTTTTGTCTTATAAACAATTTGTGTGTGGAAGG 1120
AAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTACCTCTTTCGGAAAATGATGAA 1190
TAAASATAGAGCCGATGACACTGGCTGCTAGTAGTATGAGTGTAGAATTGCTTTTTCATCGTCTCAACCT 1260
CCGCAATGAGCTTCCCTCCCTCTCATCACTGACAAATTAATGTGGGTTTATGCGCTCTTTCTCTATTCCG 1330
CCACCTATTCTGGGTTACCACAAACTGGAATACATTTTACACTATTCAAGCCATTTATTTTGAATTTA 1400
1410 1420 1430 1440 1450 1460 1470
ATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAAACGATAATTTGATTTTAAA 1470
AAATCTGAAAAGTTTCAAAAAATCAATTAATATTCCCTAACAAATGTATGGCTAAATTTTATTTCTGAC 1540
TGTTCACAAATATTTATATGTATCACTGTTTTCCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAAG 1610
CTCCGTCTCACATTTCCCAATGCTATGATCGCTTACACACACATATCCAAAAATTAAGCTAGACGGTGA 1680
TAATTTATTGGGCACGCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAGGCACAGATTATCTATCAAA 1750
1760 1770 1780 1790 1800 1810 1820
TTCAAATTTTGAACATTCAGCCAGTTCTGTAACAATTTTCCATGCTTTTTCGCCCATTAATAAACTTTCTCA 1820
CCCTTTTCATCCATCTCACTCGTATCATAAAAAGTATAGCAAAAGCCCGACCTGACCTTTTAAAGAGAAGGA 1890
GATACTGAGCCACATGCGGTGTGACCTTTTCATCTGCTCGTTCGGTCTCAAATTCACGCTCATACTAA 1960
CTCTTCAAATAGCCATAGACCTCCTTGTTTCTCTCTGCTTTTGAAGCTCGCGCCATTTTGTGTTGGTGC 2030
TGAAGCCCGGGAATAATAGTATATTTAGAGCTTATCTTTATGCAATACATAAAATAAGAGGCAATTTA 2100
2110 2120 2130 2140 2150 2160 2170
AAAAATTAATAATGAGGTGTGAGATGTAGATTTGAAAAGAGAAAAAAACATAACAAATAGGAAC 2170
GCGCAGATCAAAATTTCTATTTAAAGCTTTTCAAGATGTTTGGGCAAGATTCGGCTCAACAGAAAACGAA 2240
GTCCCTGCATAAAATGAGTGAACGTTTAGATTGAACCTCGGAAATCCTAAGCTGTAACATAGCTTTAAT 2310
CTAGATCTTAGTTGCGCTAAGCTCAAGCCCAAGCAGAAAATGACTTGCATTTAGTTTAAAGCTTAGATGA 2380
CTTGCCTGCTTCACTCTAATTCAGACATGATTTTCCAAAGAGAGTTTCAATTTTAAATGTTTCCATTTCT 2450

Figure 56

Monday, 1 December 1997 14:13
fig 56 pNP10

Page 2

2460 2470 2480 2490 2500 2510 2520
TGTTACTTAAATCTTAATGCCCTGGATGCGTAAATCGTTATCCCTTTCTCTCACACTTTCATTTACA 2520
GATTCATCAAAAGATGGTATCAAGCCAAAGACGTCTGGACGTAACCCACCCCATCATCAACCACGTCAT 2590
CAAAATAATACAAATTCATTCCGTCCGTCCGAGCCGTCGAGTGGCAATAATAATGTTGGCTCGACGATATC 2660
CACATCTGCCAAGAGCTTAGGTATCCGATCCTTCCGGCTTCTTTTGTAGAAATGATGATGATGATGATGAT 2730
TCATCAACGTACAGCTCTATTTGGAATCTAAACCGACCTACCTCCCAACTCCAAAACCTTCTAGACCAC 2800
2810 2820 2830 2840 2850 2860 2870
AAACCCAGCTAGTTTCGTGTTGCTACAACCTACAAAAATCGGAAGCTCAAAGCTAGAGGATCCCGGGGATGG 2870
GCCAAAGGACCCAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
AGGGTACCGGTAGAAAAAATGAGTAAAGGAGAGAACTTTTCACGAGGATGGTCCCAATGCTGATGATGAT 3010
TAGATGGTGTATGTTAATGGGCACAAATTTTCTGTGAGTGGAGAGGGGGAAGGATGATGATGATGATGATGAT 3080
ACTTACCTTAAATTTATTTGCACTACTGGAAGAACTACCTGTTCCATGGGTAAAGTTTAAACATATATATA 3150
3160 3170 3180 3190 3200 3210 3220
CTAAGTAACCTGATTATTTAAATTTTCAAGCAACACTTGTCACTACTTTCTGTTATGGTGTTCAATGCT 3220
CTCGAGATACCCAGATCATATGAAACGGCATGACTTTTCAAGAGTGCCATGCCGAAGGTTATGTACA 3290
GGAAAGAACTATATTTTCAAGATGAGGGGAACCTACAAGACACGTAAGTTTAAACAGTTCGGTACGTAAC 3360
TAACCAATACATATTTAAATTTTCAGGTCGTAAGTCAAGTTTGAAGGTGATACCTTGTTAATAGAAATCG 3430
AGTTAAAGGATTTGATTTTAAAGAAGATGGAACATTCTCGACACAAATTCGAATACAACATATACTC 3500
3510 3520 3530 3540 3550 3560 3570
ACACAATGTATACATCATGCGAGACAAACAAAAGAATGGAACTCAAAGTGGTAAAGTTTAAACCTGGACGTA 3570
CTAAGTAACGGAATATATTTAAATTTTCAGAACTTCAAAATTAGACACAAACATTGAAGATGGAAGCGTTT 3640
AAGTAGCAGACCATTTCAACAAAAATCTCCAATTGGCGATGGCCCTGTCTTTTACCAGACAAACATTA 3710
CCTGTCCACACAATCTGCCCTTTTCGAAAGAATCCCAACGAAAGAGAGACCATGCTCTTCTTGAGTTT 3780
GTAACAGCTGCTGGGATTACACATGGCATGATGAACATATACAAATAGCATTGCTAGAATTCCAACCTGAG 3850
3860 3870 3880 3890 3900 3910 3920
CGCCGGTCGCTACCAATACCAACCTGTCTGGTGCTAAAAATAATAGGGGCGCTGTCTACAGAGTAAGTT 3920
TAACTGAGTTCTACTAAGTAAGTAAATATTAATTTTCAGCATCTCGCCGCCGTCGCTGCTGACCTC 3990
TAAGTCGAATGATCTGTAAGATGCTAGATGGTGTCTTCTGGCTGTGCTGGGAGCGGATGATGATGATGAT 4060
TATCAAAAAAATCTTCTTAAATTTCTTTGTTTTCAGCTTCTTTAAGTCACCTCTAACAATGAATTTG 4130
TGTAGATTCAAAAAATGAATTAATTCGTAATAAAAAAGTCGAAAAAATGTGCTCCCTCCCCCATTAAT 4200
4210 4220 4230 4240 4250 4260 4270
AATAATTCTATCCCAAAATCTACACAATGTCTGTGTACACTTCTTATGTTTCTTACTTCTGATAAAT 4270
TTTTTTTGAACATCATAGAAAAAACCGCACAAAAATACCTTATCATATGTTACGTTTTCAGTTTATGAC 4340
CGCAATTTTTTATTTCTTCCGACGTCGGGCTCTCATGACGTCAAATCAATGCTCAATGCTGAAAGTTTT 4410
GGAGTATTTTGGAAATTTTCAATCAAGTGAAGTTTATGAATTAATTTCTGCTTTTGGCTTTTGGG 4480
GGTTTCCCTATGTTTGTCAAGAGTTTCGAGGACGGCTTTTCTTGCTAAAAATCAAGATATGATGAT 4550
4560 4570 4580 4590 4600 4610 4620
GCACGATGCAAGAAATATCGGAAGAAGGTTTGGGTGTGAGGCTCAATGGAAGGATGATGATGATGATGAT 4620
TTTGAAGTGGAGTAGTGTCTATGGGTTTTTGCCTTAAATGACAGAAATCATTCCTAATATACCAAAACA 4690
TAACGTTTCTTACTAGTGGGCTGACGGGCTTTCTGCTCTCGGCGCTTTCTGCTGATGACGGTGAAGAAC 4760
TCTGACACATGCAAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCGGGGAGCAGACAAAGCCG 4830
TCAGGGGCGGTCAGCGGGTGTGTTGGCGGTGGCTTAACTATGCGGCATCAGAGCAGATGATGAT 4900

Monday, 1 December 1997 14:13

Page 3

fig pNP10

4910 4920 4930 4940 4950 4960 4970
CTGAGAGTGCACCAATGCGGTGTAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGGC 4970
CTTAAGGGCCTCGTAACGCCATTTTATAGGTAAATGTCATGATAATAATGGTTTCTTAGACGTCAG 5040
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAAACATTCAAATATGTA 5110
TCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA 5180
CAFTTCCGTGTCGCCCTTATTCCTCTTTTGGGGCATTTTGCCTTCCTGTTTTTGTCTACCCAGAAACGC 5250
5260 5270 5280 5290 5300 5310 5320
TGGTAAAGTAAAGATGCTGAAGAFCAGTGGGTCACGAGTGGGTTACATCGAACGGATCTCAACAG 5320
CGGTAAAGATCCTTGAGAGTTTTCGCCCGGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA 5390
TGTGGCGCGGTATTATCCCGTATTGACGCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGA 5460
ATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG 5530
CAGTGC TGGCATAACCAAGAGTGAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG 5600
5610 5620 5630 5640 5650 5660 5670
GAGCTAACCGCTTTTTCGACAACTGCGGGGATCATGTAACCTCGCTTGATCGTTGGGAACCGGAGCTGA 5670
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCC TGTAGCAA TGGCAACAACGTTGCGCAACT 5740
ATTAACCTGGCGAACTACTTACTTATGCTTCCCGGCAACAATTAATAGACTGGATGGAGCGGATAAAGTT 5810
GCAGGACCACTTCTGCGCTCGGCGCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCGGGTGAGC 5880
GTGGGTC TCGCGGTATCAT TGCAGCAC TGGGGCCAGATGGTAAAGCCCTCCCGTATCGTAGTATCTACAC 5950
5960 5970 5980 5990 6000 6010 6020
GAGGGGGAGTCAGGCAACTATGAGTGAACGAAATAGACAGATCGCTGAGATAGGTGCTTCACTGATTAA 6020
CAT TGGTAAC TGTACAGCAAG TTTAC TCA TATATACTTTAGATTGATTTAAAAC TCAATTTTAA TTTA 6090
AAAGGATCTAGG TGAAGA TCTTTT TGAATACTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA 6160
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTC TCGCGGTAACTCTGC 6230
TGCTTCCAAACAAAAAACCCCGCTACCAGCGGTGGTTTGT TGGCGGATCAAGAGCTACCAAC TCTTT 6300
6310 6320 6330 6340 6350 6360 6370
TCCGAAGGTAAC TGGCTTCAGCAGAGCGGAGATACCAATACTGTCTTCTAG TGTAGCGGTAGTTAGG 6370
CCACCCTTCAAGAACTCTGTAGCACCGCTTACATACC TGGCTCTGCTTAA TCC TGT TACCAGTGGCTGCT 6440
GCCAGTGCCGATAAGTCGTGTCTTACCGGG TGGAC TCAAGACGATAGTTACCGGATAAGGCGCAGCGGT 6510
CGGGCTGAACCGGGGGTTTGTGTCACATACCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT 6580
ACAGCGTGAGCATTGAGAAAGTCCACCTTCCCGAAGGGAGAAAGCGGACAGGTATCCGGTAAAGCGC 6650
6660 6670 6680 6690 6700 6710 6720
AGGGTCGGAAACAGGAGAGCGCACGATSSGAGCTTCCAGGGGGAAACGCC TGGTAICTTTATAGTCTCTGC 6720
GGTTTCGCGACCTCTGACTTGAGCTCGATTTTGTGATGCTCGTCAGGGGGGGGAGCCCTATGGAAAA 6790
CGCCAGCACTGCGGCTTTTACGG TCTTGGCTTTTTC TGGCTTTTGTCTACATGTTCTTTCTTGC 6860
TTATCCCTGATCTG TGGATAACCGTATTACCGCT TGTAGTGAGCTGATACCGCTCGCCGACGCCGAA 6930
CGACCGAGCGCAGCGAGTCACTGAGTGGAGGAGGAAGCGGGAAGAGCGGCCAATACGCAAAACCGCTCTCCCGC 7000
7010 7020 7030 7040 7050 7060 7070
GGGTTCGCCGATTCATTAATGCAAGCTGGACGACAGCTTCCCGATCGGAAAGCGGCGACTGACCGCAAC 7070
GCAATTAATGTGAGT TTAGCTCACTCATTAGGCACCCAGGCTTTACAC TTTATGCTTCCCGCTCGTATGT 7140
TG TGTGGAATTGTGAGCGGA TACAATTTACACAGGAACAGCTATGACCATGAT TACGCCAAGCTGTA 7210
AGTTTAAACATGATCTTACTAAC TAACTATTCTCA TTTAAATTTTCAAGCTTAAATAAGGTGGAATCA 7280
CTCACAACGATGGATACGCTAACAAT TGGAAATGAAAT 7319

Figure 57

Monday, 1 December 1997 14:12

ig PCB501

Page /

10 20 30 40 50 60 70

ATACCA TGA TACGCCAAGCTTGCATGCTGCAGGAATTGCATATCAAGCTTATCGATACCGTCGACCT 70
GAGGATCAGAAAGAAATTGGACCAACTACCCACATCCA TTAAGCCACCGCGGTTTC TAAAGTGAGTTTAA 140
TTTTGAGTTTACGAC TACAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCATTCTGTATGACT 210
AGTTGTTGAGTGATTTTTCATTGAGAAAATATTTAAAGGAACATTTATTTACTTTTGC TTAATTGCCCCAAC 280
TTTGATTTAGTTTTCGATCAACTAGATCTTACAAAACCTTGAATACAATTCATTTTTCAGATTACCCCTC 350

360 370 380 390 400 410 420

GCACGTGTCGCCACGTGAGCAACCGCTTCAGCAAC TACCCAAA TCCAACTTTCACAAAATGTCAACA 420
TCCAGGCTTCGAGACTCCACAGTCAAGAATATCGAAAAT TGGTAAGAA TTTATTTTGAGCTCAAACTTTG 490
ATAAAATGCCCAGAAAAGAAGATGATAAAAATGTAGTTT TTTTGCAAAACCTCCACCTTTATTGCTCTAA 560
TATGACGGCTTATATCTCAATTTTCTTGAATTTATCAAAAAATTTTCACATATACAAAATG TAAAGAAAG 630
ATTTTGCACAAA TTTTGTGAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTTAA 700

710 720 730 740 750 760 770

AGTGAAGGAGGCGCAAGTCTATAC TGGAAATTAATGATCTGAAACAAAATTTGTGCTATTCTCAAAATGTTA 770
AGACATGTTTTGAAGATTTTTCAAATTCGACCTAGTTTCAGAACCTTCTTTTGTATGAAAAAGTAAA 840
AAAAAAC TATTTCAAAACCTCACCAGCCACCATGTTTCAACTCTTAATTTTATAAAATTTTGCATTTAC 910
AAATCGGC TCCCTTGGCCGAAAAGTGGCCACCAAAATCAATTTCTCGGCTTCATAATGAC TTTTAAATTT 980
TATG TGAAGAAACACAGAAAGAGGCTAACTAAATTCACAGGGACAGGTTGTCCCTCTTTCGCCCTCCCTTC 1050

1060 1070 1080 1090 1100 1110 1120

CCGCTCCTCCTCCCTCCCTTCCATCTCCAACAACAACATTTTCCAATTTCTGTTGTCCTATTTGCTTAATAA 1120
TATTTGTG TGGAAAGAACTACACGGGAGACGGTCAATTAATTCGAA TGAAGAGCA TGGCAAT TACTTC 1190
TTTCGGAAATTTGATGAATAAAGATAGACCGGA TACACTGGCTGGTAGTAGTATGAGTGTAGAAATTCCTT 1260
TTTCATCGTCTCAACTTGGSCA TGAATTTTCCCGCGCTCTCATCACTGACAATTAATGTCCGGTTTTATG 1330
CGCTCTTTCCCTATTCGGCCACTCATCTCGGTTACCAACAACTGGAATACATTTTACTACTAT TCAAGCC 1400

1410 1420 1430 1440 1450 1460 1470

ATTTATTTTGATATTTAA TTTTGTGCAATTAGGGATAAACACGACTTTTAAAGTTTATTTAAAAAAACG 1470
ATATTTTCGATTTTAAAAAA TCGAAAAGTTTCAAAAAATCAATAAATA TGGCAACAAATTTGTATGCC 1540
TAAAAATTTATTTCTACTGTTGACAAATATCTTTATATGATCACTGTTTCCATCTCAAAACCTTGAAATC 1610
CCCCAAGTTATAGGAAGCTCCGTGTACATTTCCCATGCTATGAA TCGCTACTCAGCACATATCCAAAAA 1680
TTAAGCTAGACGGTTGATAATTA TTTGGGCACGGTAATAAAGTGGCAAGCAG TTAGAATTTTAATTCAAGC 1750

1760 1770 1780 1790 1800 1810 1820

ACAGATTATCTATCAAAATTCAA TCTTGAACATTCACCCAGTTCTGTACAA TTTCCATGCTTTTGGGCC 1820
ATTAAAAAATCTTTCACCTCTTCATCCATCTCACTCG TATCATAAAAAGTATAGCAAAAGCCCGACTCT 1890
ACTTTTAAAGAGAAGGAGATCTGAGGCACATGGCG TGTACCCCTTTTCATCTCGTCCGTTCGGTCTCA 1960
ATTACGCTCA TACAACTTTCATATAGCCTAGACCTCCCTTTGTTTTCTTCGTTTGTACTCCGCC 2030
TATTTTGTGGCTGCTTGAAGCCGGGAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100

2110 2120 2130 2140 2150 2160 2170

AAAAACGAGGCAATTTAAAAATATTAATTAATTAAGGTTTGTAGATGTAGATTTTGGAAAAGAGAAAAAA 2170
ATAAAAACAAATAGGAACCGCCAGATCAAAATTTCTATT TAAAGGTTTTCAGATGTTTAGGCAAGATTCCG 2240
CTGAACAGAAAAC TGAAGTGCC TGCATFAAATCTAGTGTAACTTTAGATTGAACCTGGGAAA TCC TAAAGCC 2310
TGAACATAGCCTTATCTTATGATCTTATTTGGCA TAAAGCTCAAGCCCAAGCAGAAATGAC TTTCAATTTA 2380
GTTTAAAGCC TGAATGACTTCTTGTCTCAGTCTAATCCAGACTAGATTTCCAAGAGATTTTCAATTTT 2450

Monday, 1 December 1997 14:12

Page 2

flg pCB501

2460 2470 2480 2490 2500 2510 2520
AAATGTTTCCAGTTTCCTGTTACCTAAAATCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTC 2520
TCACACCTTCAATTACAGATTTCATCAAGATTGGTATCAAGCCAAAGACGTCCTGACCTTAAACCACTTC 2590
ATCATCAACCACCTTCATCAAAATAAATAAAATTCATTCCGTCCTCGAGCCCTTCGAGTGGCAATATAAT 2660
GTTCGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTATCCGATCCTTCGGCTTCCTTTTAGAAAT 2730
ATATTATTTTCAGAAATCATCATCAACGTACAGCCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800
2810 2820 2830 2840 2850 2860 2870
AAAACTCTCTAGACCACAAACCCAGCTAGTTCTGTGTGCTACAACACAAAAATCGGAAGCTCAAAAGCTA 2870
GCCGCTCCGAAAGCCGTGAGCACCCCAAACTTGCTCTCTGTAAGACCTATGAGGAAAACAGAGCCCG 2940
ATAACAGCGGTGGTGGTGGTGGTGGTGAATGCTGAAATTAAGTTATTCAGTAGCAAAAACCACTCTCTCT 3010
ATCGAATAGCCCAACAACTACGAGAAAGGCGGGCGGCTGCTCAACAACAACTTTGTCGAAAATCGCT 3080
GCCCGAGTGAAGAAGTGGCTGAAGCCCGGACCAAGTAAAGTGGGAAGTGGCAGCTCTATGTCGAAGCTTT 3150
3160 3170 3180 3190 3200 3210 3220
GTACGCCAAAAGTTTCTACCTAAAACGGACGCCCAATCATATCTCAACAAGACTCGAAACGATGCTC 3220
AAAGAGCAGTGAAGAAGAGTCCGGATACCTGCTTCAACAGCACGTCCGCAACGTCTATCATCGACGGA 3290
GGTTCCCTAAGCATGCTTCCACATCTTCCAAAGAGTTCAAGCTCAGACGAAAAGTCTCCGTCTATCAGAC 3360
ATCTTACTCTTAACGCTCTCATCTGTGACAGCTATCAGACACCGGATAGCCGCAACACCGGTTTCTCCAAA 3430
TATTATCAACAGCTCTTTGAGGAAAAACCAACACCTGGCAGTGAAGGAGTGAAGAGCACAGCGAAAAAA 3500
3510 3520 3530 3540 3550 3560 3570
GATCCACCTCCAGCTGTTCCGCGACCTGACACCCAGCCAACTCCGAGTGTCTAGTCCAATTATGGCAC 3570
ATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACAGAACTGAAAAGCTCCCAATCAATGAGCAT 3640
CGACACGACGGACGTTCCACCGCTTCCACCTGTAAGAACTAGTTGTTCCACTTAAAAAGACTTCAATCCGA 3710
CAACCAACCAACGTACGATGTTCTTCTAAAACAGGAAAAATCACATCGCTGTCAAGTCTGTTGGATATG 3780
AGCAGTCTGTCGGCTCTGAAGACTCCATTGTGGCTCATGCTCGGCTCAGGTGACTCCGCGACAAAAAC 3850
3860 3870 3880 3890 3900 3910 3920
TTCTGCTAATCATTCGCTCGAGAGAAAGGATGGGAAAGAATAAGACATCAGAAATCCAGCGGCTACACCTCT 3920
GACGCGGGGCTGCGGATGTCGCGCAAAATGAGGGAGAAAGCTGAAAGAAATAGGATGACATGACCTGTCGAG 3990
CACAGAACGGCTATCTGACAACTTCGAAGACAGTTCCTCTTGTCTGCTGGAATATCCGATAAACACGA 4060
GCTCGACGACATATCCACGGACGATTGTCTCGGAGTAGACATGGCAACAGTCTGCTTCAAAACAAGCGAC 4130
TATTCCTACTTTGTCTGCGCACTCCACGCTCTCTCTCTCAAAGCCCCGAGTCCCGAGTGGTCTCTCATAT 4200
4210 4220 4230 4240 4250 4260 4270
CAGTCCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAACTCTCTGTCCTCAGTGGCGAACGAGCCA 4270
ACCTGGCGCGCTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCGGGATCTCATCTATTCT 4340
CCACACTTATCAGTGTCTAGCTGATAAGGACACAATGCTATCCACTCACAGACTAGTCTGACGACCTCT 4410
CACAAAAACCAAGCTATTCAGGCAATTTCTATCTGATCTGTAATGCCACCTTCAAGAGTTCACATC 4480
CACCGAGCACAGAATGGCGGCTCTCTTGAAGCCCGAGACGGGTCGGAACCTGATGTGAAAATATGATTCT 4550
4560 4570 4580 4590 4600 4610 4620
TCAGGATCTTACTCGGCGCGTCTCCGAGGTGGAAGCTCTACTGGTATCTTATGAGAGAGAGCTTCCAACTCT 4620
ACAGACTATCCGATGAAAAATCCCGCGCACATCTGCCAAAAGTGAGATGGGAATCCAACTATCACTGCG 4690
TAGCAGGACAGCATATGGAATCTCTCAATGAGAAGTACGAACAATCTATTCGGGACATGGCAGTCACTTG 4760
GAGTGTACAAAGAACATGTCGACTCACTAAGCAAGAACAGGAGAACATGGAAGCATGTTGATCTCT 4830
TTGAGCAAAAGCTTAGAAAACTCACTCAACACATCTGATCGATCCAACCTTGAAGCTTGAAGAGGCAATACG 4900

Monday, 1 December 1997 14:12

Page 3

flg pCB501

4910 4920 4930 4940 4950 4960 4970
A"TCAGGCAGGACATTGCTCATTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAAC 4970
GAAGGCGCTGGTGAGCTTCTTCTGCAACCACTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCAT 5040
CGTCGTCGAAAAGCAGCAAGCAGGAGAAATCAGCTTGAGCTCGTTTGGCAAGAACAAGAAGAGCTGGAT 5110
CCGCTCTCACTCTCCAAGTTACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATT 5180
TCCGGATCTCAAGGAACCTTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGGGATA 5250
5260 5270 5280 5290 5300 5310 5320
GTGCACTTTACGAAGTCCGCTTTGACAATCTGGATCGTSCCCGCGAAGTTGATGTTCTGAGGGAGACAGT 5320
GAALAAATTTGAAAALLGAGAACTAAGCAATTAAAGAAAGAGTGGACAAACTCACCACGGTCCAGCCACT 5390
CGTGCTTCTTCCCGCGCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCA 5460
GTACATCAGCTAGTCAATCTTCGAAACGATCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGA 5530
CAICCGCTGGAGAAATCAGTTCCGATCGTTAACGGGGACTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGC 5600
5610 5620 5630 5640 5650 5660 5670
AAGGTCAGTGAAGAAATTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACATAATTT 5670
CTAAAATGGACCCAGCTCTACCTCGGACTAAGCACTGAGTCCAATGGCTACAGCATCAGCCAGCT 5740
GAAACGAGTGTGGATGCAGAGCCCCCGAGATGCCCTCTTGGCGTGGAGGTGCAATAACATATCAGTC 5810
TCCCTCAAAGGTCTGAAGGAGAAAATGCTCGACAGCTGCTGTTCCGAGACGCTGATCCCCAAGCCGATGA 5880
TGCAGCACTACATAAGCTCTCTGTAAGCAACGGGCGCTCTCTCTCGGCCCCAGCGGCACGGGCA 5950
5960 5970 5980 5990 6000 6010 6020
GACCTACCTGACCAATCGCTTGGCCGAGTACCCTGGTGGAGCGCTCTGGCCGTGAGGTCAAGAGGGGATC 6020
GTCAGCACTTCAACATGCAACAGCAGTCTTGAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAAC 6090
AGATAGACCCGGGAAACAGGAATTGGGGAATGCCCCCTGGTATCTATTGGATGACCAGAGTGAAGCAGG 6160
CTCCATCAGTGAGTTGGTCAATGGGCCCCCTCACTGCAAGTATCATAAATGTCCCTATATTAAGGTAAC 6230
ACCAATCAGCTTGAATAATGACACCCCAACCATGGCTGCACTTGAAGTTCAAGGATGTTGACCTTCTCCA 6300
6310 6320 6330 6340 6350 6360 6370
ACAACGTGGAGCCAGCCAATGGCTTCTGTCTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGA 6370
CATCAATGCCAACAAGGAAGAGCTGCTTGGGGTGTCTGACTGGGTACCCAAGCTGTGGTATCATCTCCAC 6440
AGCTTCTCTTGAAGAGGAGAGGAGGTCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6510
GCATGAGGACTTCCGGACCTGGTTCATGAGCTGTGGAACAACCTCTATCATTCCTTACAGGAAGG 6580
AGCCAAGGATGGGATAAAGTCCATGAGACAGAAAGCTGCTTGGGAGGACCCAGTGAATGGGTCCGGGAC 6650
6660 6670 6680 6690 6700 6710 6720
ACACTTCCCTGGCCATCAGCCCAACAGACCAATCAAGGCTGTACCACCTGCCGCCACCCACCGTGGGCC 6720
CTCAGCATTTCCCTCACCCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACACAGA 6790
TCTCTGATGGCCATGCTGCTGAACCTTCAAGAAGCTGCCAATACATGAGTCTCCAGATCGAGAAACC 6860
ATCTGAGACCCCAACCTTCAAGCAACACTTGAAGGTTCCGCAATCACTGTCACCCCGGACAGCAGAAC 6930
GCTGGCATCAGCTATCTTAGCT 7000
7010 7020 7030 7040 7050 7060 7070
AGGAGAACAGGAGGAGGAGGAGATGAAAGAGGAGGACAGGTTCTTGGTCTGTACCTTTGAGAACCTCT 7070
CTAGGAACGAAATGGTGGGGTGGGCTTTGGGAACCTTGTGCCCCCAAACACATTTACTGGCCCTCTCTAAT 7140
GACTTTGGGGAAAAGATGATTCGGGTCTTTCCCTTGACCTCTTGTCTTCAATTACAACTCTCTGGCTTT 7210
CTGGGGAGGGGTTCAGAAAACATCAAAACACTGCAAGCAGTTCCCCGGAATTCAGCTTGGACCTTAACAGG 7280
CTGAACCTGCTCAAAAGAAGCCGAATTCCAGCAACTGGCTCCCCATGGTATGATTAATCTGAGCTCCGC 7350

Monday, 1 December 1997 14:12
fig PCB501

Page 4

7360 7370 7380 7390 7400 7410 7420
ATCGGCGCGC TGT CAT CAG ATCGCC ATCTCGCGCCCGTGCCTCTGACTTCFAAGGCCAAATACGCTTCAAC 7420
ATCCCTACATGCTCTTCCTCCCGTGGTCCACCCCTATTTTGTATTATCAAAAAACCTCTCTTA 7430
ATTCCTTCTGTTTTTACGTTCTTTTAAGTCACCTCTAACAATGAAATTGTGTAGATCAAAAAATAGAATT 7440
AATTCGTAATAAAAAGTCGAAAAAAATTTGTGCTCCCTCCCGGCTTAATAATAATTTCTATCCCAAAATCT 7450
ACACAATGTCTGTGTACATTCCTTAATGTTTTTTTACTTCTGATAAATTTTTTTTGAACATCATAGAA 7460
7710 7720 7730 7740 7750 7760 7770
AAAACCGCACACAAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTATTTCTTCGCA 7770
CGTCTGGGCGCTCTCATGACGCTAAATCATGCTCATCGTGAAAAAGTTTTTGGAGTATTTTGGAAATTTTC 7780
AATCAAGTGAAGTTTATGAAATTAATTTCTGCTTTTGTCTTTTGGGGTTTCCCTATTTGTTGTCA 7790
AGAGTTTCGAGGACGGCTTTTCTTGTCTAAAATCACAAGTATTGATGAGCAGGATCAAGAAAGATCGG 7800
AAGAAGTTTGGGTTTGAGGCTCAGTGAAGGTGAGTGAAGTTGATAATTTGAAATTTGAGTGTCT 7810
8060 8070 8080 8090 8100 8110 8120
ATGGGGTTTTTGCCTTAAATGACAGAATACATTCCCAATATACCAAAACATAACTGTTTCTACTAGTGG 8120
CGGTACGGGCGCTTTCTGCTCGCGCGTTTGGTGTATGACGCTGAAAAACCTCTGACACATGACGCTCCCG 8130
AGACGGTCACAGCTTGCTGCTGAAGCGGATGCGGGGAGCAGACAAGCCGTCAGGGC3CGTCAAGCGGTTGT 8140
TGGCGGGTGTGCGGGCTGGCTTAACATATCGGCTACAGAGCAGATTGTACTGAGAGTGCACCAATATCGG 8150
TGTGAATATCCGCACAGATGCGTAAAGGAGAAAAATACCGCACTAGGCGGCTTGAAGGCGCTGCTGATAC 8160
8410 8420 8430 8440 8450 8460 8470
CTATTTTATAGGTTAATGTCTATGATAATAATGGTTTTCTTAGACGTCAGGTGGCAGTTTCTGGGGAATG 8470
TGGCGGAACCCCTATTTGTTTATTTTCTAAATACATTCAAAATATGATCCGCTCATGAGACAAATACC 8480
CTGATAAATGCTTCAATAATATGAAAAAGGAAGAGTATGAGTATCAACATTTCCCTGTGCGGCTTATT 8490
CGCTTTTTCGGGCTTTTGCCTTCTGTTTGTCTCACCAGAAACGCTGTTGAAAGTAAAGATGCTG 8500
AAGATCAGTTGGGTGCACGAGTGGTTACATCGAACTGATCTCAACAGCGCTAAGATCCTTGAGAGTTT 8510
8760 8770 8780 8790 8800 8810 8820
TCGCCCCGAAGAAGCTTTTCCAAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGGGTAATACCGG 8820
ATGACGCGGGGCAAGAGCAACTCGGTCCCGCATACATATTCAGAGATGACTTGGTTGAGTACTCAC 8830
CAGTCACAGAAAACCATCTTACGGATGCAATGACAGTAAAGAGATTTATGCAGTGTGCCATTAACCATGAG 8840
TGATAACACTCGCGCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAAGCGCTTTTTCGAC 8850
AACATGGGGGATCATGTAACCTGCTTGGTATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAACGAGC 8860
9110 9120 9130 9140 9150 9160 9170
AGGCTGACACCAGGATGCTTGAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACGATTTAC 9170
TCAGCTTCCCGGCAACAATTAATGAGCTGGATGGAGGGGAGTAAAGTTGAGGACCACTTCTGCGCTCG 9180
GCGCTTCCCGCTGGCTGCTTTATTGCTGATAAATCAGGAGCGGTTGAGCGTGGCTCTGCGGTAATATG 9190
GAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCACTAT 9200
GGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCAGTATTAAGCATTTGTAAGTCTGAGACCAA 9210
9460 9470 9480 9490 9500 9510 9520
GTTTACTCATATATACTTTAGATTTTAAACTTCAATTTTAAATTTAAAGGATCTAGGTAAGATCC 9520
TTTTTGAATCTCATGACCAAAAACCTTAAAGTGTAGTTTTCCTTCCACTGAGCGTCAGACCCCGTAGA 9530
AAAGAATCAAGGATCTTCTTGAGATCTTTTTCGCGGTAATCTGCTGCTTGCACCAAAAAAACCA 9540
CGGCTACCAAGGTTGTTGTTTGGCGGATCAAGAGCTACCAACTCTTTTCCCAAGGTAATCTGCTTCA 9550
GCAGAGCGCAGATACCAAACTGCTGCTTAGTGTAGGCTAGTTAGGCCACCACTTCAAGAACTCTGT 9560

Monday, 1 December 1997 14:12
fig pCB501

Page 5

```

          9810      9820      9830      9840      9850      9860      9870
+-----+-----+-----+-----+-----+-----+-----+
AGCACCGCCTACATACC TCGCTCTGCTAATCCTGT TACCAGTGGCTGCTGCCAGTG GCGATAAGTCGTGT 9870
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCG 9940
GCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT GAGCATTGAGAAAG 10010
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGC 10080
ACGAGGGGAGCTTCCAGGGGGAAACGCC TGGTATCTTTATAGTCCTGTTCGGGTTTCGCCACC TC TGA CTTG 10150
          10160      10170      10180      10190      10200      10210      10220
+-----+-----+-----+-----+-----+-----+-----+
AGCGTCGATTTTTG TGA TGCTCGTCAAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGGCC TTTT 10220
ACGGTTCCTGGCC TTTT GCTGCCCTTTTGCTCACA TGTTCCTTCCTGCGTTATCCCCTGATTC TGTGGAT 10290
AACCGTATTACCGCC TTTT GAGT GAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAG 10360
TGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCGCCTCTCCCGCGCGTTGGCCGATTCAATTAATG 10430
CAGCTGGCAGCAGAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC 10500
          10510      10520      10530      10540      10550      10560      10570
+-----+-----+-----+-----+-----+-----+-----+
ACTCATTAGGCACCCCAAGC TTTACAC TTTATGCTTCCGGCTCGTATGTTGTGTGGAA TTTG TGA CCGGA T 10570
AACCA TTTACACAGGAAACAGCT 10594

```

Friday, 28 November 1997 11:36
pTB115

Fig 58

Page

Created: Friday, 28 November 1997 11:02

10 20 30 40 50 60 70
ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTctagAAATGFAAACCTGTCATTTCTGTG 70
TATTTACAGCACCGCAGAGAGCACCATAAACAGCTACAACAAGAGCACGAACGTCGCTCCAACAGAAAAA 140
GC!GCAAAACGC TGGTACTCACCACATGCCTGTGATAATTGCCATTTCCTTCGATTCTIGACTTTCAAATTG 210
TGAATATGATAACCTAAAATCTGCCATTCAAAATGAAAACCTTAAAGTTTAAAGAGCCTTCAACGACCC 280
CGCGACCTTCCCACCTCATCGTTTTTCCCTTCTATCTTATTCTAAATTTTTTGTATTTTGTGTCAGTTTGC 350
360 370 380 390 400 410 420
TATATCC!CATCACCCTTTTACCGTTTAAATCTCTCTATAGTATTGTTCCGGTGGTTTCAAGATGAAACG 420
TTTTGTGTGATCATTTTTGAACACAGCGxGAAAAATGCAGAAACATCATCCACAAATGCATTAAGTGC 490
ACTCATTGTGATGACTCCTAACCCCTCGCCCCACCATTGTCTCTTCACAAATCA!GGCACAATTCAA 560
AAATGGCTTTTTGTGTTGGATCTTGATCTCGTCCTTTCTCTCTTTTCGTTCTCTCGGACCACTCAT 630
TATTCGAATGCCATGCATTGTGTGATGATGCGCAGATGGTCGCACACACTACAACAGATATGATGGTGG 700
710 720 730 740 750 760 770
CTGTGCGTGGTGCAGAGCTCATGAATAAGTGATGGGGCACCGAGATGTCGACTCCCTCCATTATTCCTCTT 770
GAGTCGCATCTTTGTCTTGGTCGGTGTGTCAGTCGAGCCTCAAGGCAGCCGTTCCGCCAGTGATCGCCTG 840
CTAAGAAGTTGTAGGTGTAGAGGAAGAAACGTCGGGTCCAAATTTCAAGCAACATCCGAGATCGTGA 910
tcccgGGGATTGGCCAAAGGACCCAAACGTATGTTTCGAATGATACTAACATAACAAGAAACATTTTCAG 980
GAGGACCCTTGGctagactagtgatccgagctctcccatATGACGACGTCAAAATGTAGAATTCGATACC 1050
1060 1070 1080 1090 1100 1110 1120
AATCTACACGATTGGGCCAATCGGCACCTTTTCAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTC 1120
AATGATTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTCGATCTTCGATCAACGAATTCCTCC 1190
CTGCATTACGAAACGTTGGCAAAAAACCATCGAACCTGGATGGCCTCGAAACGTGTCCTGACTACCT 1260
GAAALATCTGGGTCTCGACTGCTCGAACTCACCAAAACCGATATCGACAGCGGAAACCTGGGTGCAGTT 1330
CTCCAGCTGCTCTTCTGCTCTCCACCACAAAGCAGAAGCTTCGGCAACTGAAAAAAGATCAGAAGAAAT 1400
1410 1420 1430 1440 1450 1460 1470
TGGAGCAACTACCCACATCCATATGCCACCCGCGGTTTCTAAATTACCTTCGCCACGTGTGCGCACGTC 1470
AGCAACCCGCTTCAGCAACTAACCCAAATTCCAACTTTCACAAATGTCAACATCCAGGCTTCAGACTCCA 1540
CAGTCAAGAAATATCGAAATTTGATTTCATCAAAAGATTGGTATCAAGCCAAAGACGCTTGGACTTAAACCAC 1610
CCTCATCATCAACCACTTCATCAAAATAACAAATTCATTCCTGTCCTCGAGCCGCTTCGAGTGGCAATTA 1680
TAATGTTGGCTCGACGATATCCACATCTCGGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCG 1750
1760 1770 1780 1790 1800 1810 1820
AATCTAAACCGACCTACCTCCCAACCTCAAAACCTTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTTA 1820
CAACGACAAATTCGGAAGCTCAAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCTTAAACCTGCTTCGT 1890
GAAGACTATTGGAGCAAAACAACAGCCCTAFAACAGCGGTGTTGGTGGTGGTGGAAATGCTGAAATTAAG 1960
TATTTCAGTAGCAAAACCCATCTTCTCATCGAATAGCCCAACCTACGAGAAAGCGGCGCGGCTGC 2030
CTCAACAACAAACCTTTGTGCAAAATCGCTGCCCCAGTGAAAAGTGCCCTGAAGCCGCGGACCAAGTAACT 2100
2110 2120 2130 2140 2150 2160 2170
GGGAAGTGCCACGTCATGTGCAAGCTTTGTACGCCAAAAGTTTCTACCGTAAACCGGACGCCCCAACT 2170
ATATCTCAACAAGACTCGAAACGATGCTCAAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCALCA 2240
GCAGCTCGCCAAACGTCATATCGACGGAAGGTTCCCTAAGCAATGCATTCCACATCTTCCAAGAGTTCALC 2310
GTACAGACGAAAGCTCTCCTCATCACACGATCTTACCTTTAACGCTCCATCGTGACAGCTATCAGACAG 2380
CGGATAGCCGCAACACCGGTTTCTCAAAATATATCAACAAGCCGTTTGAGGAAACACCAACACTGGCAG 2450

Friday, 28 November 1997 11:36
pTB115

Page

2460 2470 2480 2490 2500 2510 2520
TGAAAGCAGTGAAAAGCACAGCGAAAAAAGATCCACC TCCAGCTGTTCCGCCACGTGACACCCAGCCAAC 2520
AATCCGAGTTGTTAGTCCAA TATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACCA 2530
GAACCTGAAAAGCTCCAATCAATGAGCATCGACACAGCGGCTATCCTGACAAC TCGAAGACAGTTCCCTCC 2540
TTGTCCCTCTGGAATATCGATTAACACGACACCAACGTACGATGTTCTTCTAAAACAAGGAAAAAT 2550
CAGATCGCC TGTCAAGTCGTTTGGATATGAGCAGTCCTCCGCTCTGAAGACTCCATTGTGGCTCATGCC 2560
2810 2820 2830 2840 2850 2860 2870
TCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTCGCTGGAGAGAAAGGA TGGGAAAGAATA 2870
AGACATCAGAATCCAGCGGCTACACCTCTGACGCGGGTGT TGGCA TGTGGCCAAATGAGGGAGAAAGCT 2880
GAAAGAAATACGATGACATGAC TGTGAGGACAGAGAAGGCTATCCTGACAAC TCGAAGACAGTTCCCTCC 2890
TTGTCCCTCTGGAATATCGATTAACACGACACCAACGTACGATGTTCTTCTAAAACAAGGAAAAAT 2900
TGGCAACAGTCGGCTCCAAACATAGCGAC TAT TCCCACTTTGTTCCGCATCCCACGTCTTCTTCC TCAAA 2910
3160 3170 3180 3190 3200 3210 3220
GCCCCGAGTCCCCAGTCCGCTCTCCACATCAGTCGATTCTCGATCTCGAGCAGAAACAGGAGAAATGTGTAC 3220
AAAC TCT TGTCCCACTGCGGAACGAGCCAAAG TGGCGCGGCTGCCACCTCAACCT TCGGACAAAC TCGC 3230
TAAGATCCCCGGGATAC TCA TCC TAT TCTCCACAC TATCAGTGTGAGCTGATTAAGGACACAA TGTCTAT 3240
GCACACAGACTAGTCGACGACCTTCTTCAAAAAACCAAGCTATTCAGGCCAA TTTCACTCACTTGAT 3250
CGTAAATGCCACCT TCAAGAGTTACATCCACCGAGCACAGAATGGCGGCTCTCTT GAGCCCGAGACGGG 3260
3360 3370 3380 3390 3400 3410 3420
TGCCGAATCGATGTGGAATATGATTCTT CAGGATCCTACTCGGCGGCTTCCCGAGG TGGAAAGCTCTAC 3420
TGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACATTCTGCAAAA 3430
AGTGAGATGGGATCCCAACTATCACTGGCTAGCAGCAGACATATGGATCTCTCAATGAGAAGTACGAAC 3440
ATGCTATTCCGGGACATGGCAGCTGACTTGSAGTGT TACAAGAACACTGTGACTCACTAACCAAGAAACA 3450
GGAGAATATGGAGCA TGTGTGATCTTT TGAAGCAAAAGCTTAGAAAAAC TCACTCAACCAATGATCGA 3460
3560 3570 3580 3590 3600 3610 3620
TCCAACCTTGAAGCCTGAAGAGGCAATACGATT CAGGCAGGACATTGCTCATT TGAAGGATATTAGCAATC 3620
ATCTTSCATCCAACCTCAGCTCATGCTAACGAAGCGGCTGGTGAGCTTCTTCTGTCACCATCTCTGGAATC 3630
AGTTGCATCCCA TCGATCA TCGATG TCA TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG 3640
TGGT TGGCAAGAACAAGAAGAGCTGATCCGCTCCTCACTCTCCAAGTT CACCAAGAAGAAGAACAGA 3650
ACTACGACGAAGCACATATGCCATCAATT TCCGGA TCTCAAGGAAC TCTGACAAAC TGTATGTGATTGA 3660
3660 3670 3680 3690 3700 3710 3720
TCCAACCTTGAAGCCTGAAGAGGCAATACGATT CAGGCAGGACATTGCTCATT TGAAGGATATTAGCAATC 3720
ATCTTSCATCCAACCTCAGCTCATGCTAACGAAGCGGCTGGTGAGCTTCTTCTGTCACCATCTCTGGAATC 3730
AGTTGCATCCCA TCGATCA TCGATG TCA TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG TCG 3740
TGGT TGGCAAGAACAAGAAGAGCTGATCCGCTCCTCACTCTCCAAGTT CACCAAGAAGAAGAACAGA 3750
ACTACGACGAAGCACATATGCCATCAATT TCCGGA TCTCAAGGAAC TCTGACAAAC TGTATGTGATTGA 3760
3760 3770 3780 3790 3800 3810 3820
GTTGAAGCAAGAGCTCAAAGAACCGCA TATGACAC TTTACGAAG TCCGCTTGAACATCTGGATCGTGCC 3820
CGCGAAGTTGATGTTCTGAGGGAGACAGTGAAACAAGTTGAAAAACCGAGAACAAGCAAT TAAAGAAAGAAG 3830
TGGACAAAC TACCAACCGGTCAGCCAC TCGTGCT TCT TCCGCGGCTCAATTCCAGTTATCTACGACGA 3840
TGAAGCA TGTCTATGATGACAGCTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCTCTTGGCTGC 3850
AAC TCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTGATTCGCTTAACCCGACAAAG 3860
3860 3870 3880 3890 3900 3910 3920
AGATAATCGTAGGA TATC TGGCA TGTCAACCAAGTCAGTCATGCTGGAAAGACATTGATGTTTCTATCTCT 3920
AGGACTATTTGAAG TCTACC TATCCAGAA TGTATG TGGAGCATCAACTTGGAAATCGATGCTCTGATCTCT 3930
ATCCTTGGC TATCAAA TGGTGAACCTTCCAGCGCTCATTTGGAGACTCCACAACCA TGA TAAACCAAGCAATC 3940
CAACTGACATTTCTATCTCT TCAACTACAATCCGAATGTTTCATGCACGGTGCCGCAAGAG TCGCTGAGA 3950
CAGTCTGGTCTCTGATATGCTTCTTCCAAAGCAAAATGATTCTCCAAC TCGTCAAG TCAAT TTTGACAGAG 3960

Page

BNSDOCID: <WO 9824810A2 | >

Friday, 28 November 1997 11:37

Page

pTB115

7360 7370 7380 7390 7400 7410 7420
CGGTCGCGCAACACTATTCTCAGAATGACTTGCTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG 7420
GATGGCATGACAGTAAGAGAAATTAAGCAGTGGTGGCATAACCATGAGTGATAACACTGCGGCCAACCTTAC 7430
TTCTGACAACCATCGGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACATGGGGGATCATGTAACCTG 7440
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCAACGATGCTGTA 7450
GCAATGGCAACAACGTTGCGCAAACTATTAACCTGGCGAACTACTTACTCTAGCTTCGCGCAACAATTAA 7460
7710 7720 7730 7740 7750 7760 7770
TAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTCGCTCGGCGCTTCCGCTCGCTGGTTTAT 7710
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG 7720
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG 7730
CTGAGATAGGTGCTCACTGATTAAGCATTGGTAACCTCAGACCAAGTTTACTCAATATACCTTAGAT 7740
TGATTTAAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAA 7750
8060 8070 8080 8090 8100 8110 8120
ATCCCCTAACGCTGAGTCTTCTGCTTCCACCTGAGCGCTCAGACCCCGTAGAAAAAGATCAAGGATCTTCTT 8120
ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCACCAAAAAAACACCCGCTACCAAGCGGTGGTTTGT 8130
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTACGAGAGCGCAGATACCAAACT 8140
GTCTTCTAGTGTAGCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCAGCTTACATACCTCGCT 8150
TGCTAATCTGTATACAGTGGCTGCTGCCAGTGGCGATAAGTCTGCTTACCGGGTGGAGTCAAGACG 8160
8410 8420 8430 8440 8450 8460 8470
AAGCTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGTTCTGTCACACAGCCAGCTTGGAGCGA 8470
ACGACCTACACCGAACTGAGATACCCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAA 8480
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGGAGGAGCTTCCAGGGGAA 8490
CGCCTGGTATCTTTAAGTCTTGTCTGGGTTTGGCCACCTCTGACTTGGAGCGTCGATTTTGTGATGCTCG 8500
TCAGGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCCTTTTACGGTTCTTGGCCTTTTCTGCG 8510
8760 8770 8780 8790 8800 8810 8820
CTTTTGCTCACATGTTCTTTCTTGGCTTATCCCTGATTCTGTGGATAACCGTATTAACCGCTTTGAGTG 8820
AGCTGATACCGCTCGCGCGAGCCGAACGACCGAGCGCAGCGAGTGAGCGAGGAGGCGGAAGCGGAAAGCGC 8830
CCAATACGCAAAACCGCTCTCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCAAGACAGGTTTCCCG 8840
ACCTGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAAGGACCCCAAGCT 8850
ACACCTTAAGCTTCCGGCTCGTATGTTGTGTGGAATTTGAGCGGATAACAATTTACACAGGAAACAGC 8860
9110 9120 9130 9140 9150 9160 9170
T 9101

Page

Created: Friday, 28 November 1997 10:58

BNSDOCID: <WO 9824810A2 | >

Friday, 28 November 1997 11:00
pPD95-75

Page

2460 2470 2480 2490 2500 2510 2520
TCAGTTGGGTCACGAGTGGGTTACATCGAAC TGGATCTCAACAGCGGTAAGA TCC TTGAGAGTTTTCGC 2520
CCCGAAGAACGT T TCC AATGATGAGCACTTTTAAAGT TCTGCTATGTGGCGCGS TAT FATCCCGTATTG 2530
ACGGCGGGC AAGAGCAAC TCGGTCCGCCCATACACTATTCTCAGAATGAC TCGGT T GAGTACTCACCAGT 2660
CACAGAAAAGCATCTTACGGATGGCA T GACAGTAAGAGAA TATGCAGTGTGCGATAACCATGAG TGA T 2730
AACACTGCGGCCA T TAC T T C T GACAACGATCGGAGGACCGAAGGAGCTAACCG T T T T TGCACAACA 2800

2810 2820 2830 2840 2850 2860 2870
TGGGGGATCATGTAAC TCGCC T T GATCGTTGGGAACCGGAGCTGAATGAAGCCA TACCAAACGACGAGCG 2870
TGACACCACGATGCC TGTAGCAATGGCAACAACGTTGCGCAAACTATTAAC T GCGCAACTAC T TAC T C TGA 2940
GCTTCCCGGCAACAA T T AATAGACTGGATGGAGGCGGATAAAGTTGCAAGGACCAC T TCTGCGCTCGGCC 3010
T TCCGGC TGGCTGGTTTATTGC TGA TAAATCTGGAGCCGGTGAGCG TGGCTCTCGCGGTATCAT TGCAGC 3080
ACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGGACGGGGAG T CAGGCAAC TATGGAT 3150

3160 3170 3180 3190 3200 3210 3220
GAACGAA T AGACAGATCGCTGAGATAGG T GCTCACTGATTAAGCAT TGG TAACT T GTCAGACCAAGTTT 3220
ACTCATATATACTTTAGATTGA T T TAAACTTTCATTTTAAATTTAAAAGGATCTAGTGAAGA TCC T TTT 3230
TGATAATCTCATGACCAAAA TCCCTTAACGTGAGTTTTCG T TCCACTGAGCGTCAGACCCCG T AGAAAAG 3360
ATCAAAGGATCTTCT T T GAGATCTTTTTTTCTGCGCGTAATCTGCTGC T TSCAAAACAAAAAACCACCC 3430
TACCAGCGGTGG T T T T T GCGCGATCAAGAGC TACCAACTCTTTTCCGAAGGTAACTGGCTTCAGCAG 3500

3510 3520 3530 3540 3550 3560 3570
AGCGCAGATACCAATATCTG TCTCTAGTGTAGCCGTAGTTAAGGCCACCAC T TCAAGAACTCTGTAGCA 3570
CCGCTACATACCTCGC TCTCTAATCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTA 3640
CCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGG T TCTGTGCAC 3710
ACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAAGCGCC 3780
ACGC T TCCCGAAGGGAGAAAAGCGGACAGGTATCCGG T AAGCGCCAGGETCGGAACAGGAGAGCGCACGA 3850

3860 3870 3880 3890 3900 3910 3920
GGGAGC T TCCAGGGGGAAACGCCTGGTATCTTTATAGTCTG TCGGGTTTCGCCACCTCTGAC T T GAGCG 3920
TCGA T TTTGTGATGCTCT T CAGGGGGGCGGAGCC TATGGAAAAACGCCAGCAACGCGGCC TTTTACGG 3990
TTCTGGCC T T T T GCTGGCTTTTGTCTCATATGTTCT T TCTGCGTTATCCCC T GATTCTGTGGATAACC 4060
GTATTACCCCT T T GAGTGAGCTGA TACCGCTCGCCGACCCGAAACGACCGAGCGCAGCGAGTCACTGAG 4130
CGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCGCTCTCCCGCGCGTTGGCGA T TCAATTAATGCAGC 4200

4210 4220 4230 4240 4250 4260 4270
TGGCAGCAGGTTTCCCGAC TGGAAAGCGGGCAGTGAGCCCAACCGCAATTAATGTGAGT T AGCTCAC T C 4270
AT T AGGCACCCAGGCT T TACACTTTATGCT TCCGGCTCG TATG T TGTGGAAT TGTGAGCGGATACCA 4340
ATTTACACAGGAAACAGCTATGACCA T GATTACGCCAAGCT T gtaagt ttaacatgata t t a c t a a c t a 4410
a c t a t t e a t t a a a t t t t a a g A G C T T A A A A A T G G C T G A A A T C A C T C A C A A C G A T G G A T A C G C T A A C A A 4480
CTTGGAAA TGAAT 4494

Friday, 28 November 1997 13:10
pBS KS/X16

Page

Created: Friday, 28 November 1997 12:02

10 20 30 40 50 60 70
CTAAATTGTAAGCGTTAATATTTTGTTAAATTCGCGTTAAATTTTGTTAAATCACTCATTTTAAAC 70
CAAAGGCGGAAATCGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGTTTGAGTGTGTTTC 140
CAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGICAAAGGGCGAAAAACCGTCATCA 210
GGCGGATGGCCACTACGTGAACCAICACCCIAATCAAGTTTTTGGGGTGGAGGTCCCGTAAAGCACTA 280
AATCGGAACCCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGAAAGCCGGCGAATGTGGCGAGAAAGG 350
360 370 380 390 400 410 420
AAGGGAAGAAAGCGAAGGAGCGGGCGCTAGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCAAC 420
CACACCCGCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCAATCGCCAATCAGGCAGCGCAACATGTGG 480
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCAGGGCGAAAGGGGATGTGCTGCAAGGCCAT 560
TAAGTTGGGTAAACGCCAGGGTTTTCCCAAGTCAAGCTGTGTAACACGACGGCCAGTGAAGCGCGTAATA 620
CGACTCACTATAGGGCGAATGTGAGCTCCACCGCGGTGGCGGCTCTAGAATAGTGGATCCCCCGGG 700
710 720 730 740 750 760 770
CTGCAGGAATTCGATACTAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAAGAAATTTGGAGCAACTAC 770
CCACATCCATTATGCCACCCGCGGTCTTAAGTGAGTTTAAATTTTGAGTTTACGACTACAAAAATGTGTT 840
CTTTAATAAATATCTTCGACTTGAGTCTATTCTGTATCAGTAGTTGTGAGTGTATTTTCATTGAGAAAA 910
TATTAAGGAACATATTTACTTTGCTTATTTGCCCTAACITTGATTAGTTTTTGAATCAACTAGATC 980
TTACAAAACTTGCAATACAATTCATTTTCAGATTACCTTCGCCACGTGTGCCACGTCAAGCAACCGCTT 1050
1060 1070 1080 1090 1100 1110 1120
CAGCAACTAACCCAAATTCCAACTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAAAT 1120
ATCGAAAAATGGTAAGAAATTTTATTTGAGCTCAAACCTGTATAAAATGCCCAAGAAAGAAGATGATATA 1190
AATGTAGTTTTTTTGCAAAACTTCCACCTTTATTGCTAATATGACGGCTTATATCTCAAATTTCTTTGA 1260
GTTTTATCAAAAAATTTCCACTATACAAATGTAGAAAAGTATTTGCACAAATTTTGTCAGTTGACAGC 1330
TTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTATTAAGTGGAGGAGCGCAAGTCTATACTGGAA 1400
1410 1420 1430 1440 1450 1460 1470
ATAATGATCTGAAACAAATTTGTGCTATTCCTAAATGTTAAGACATGTTTTGAAGATTTTTCCTAAATTC 1470
GCACTAGTTTCAGAACCTTCCTTTTGTATGAAAAAGTAAAAAAAACCTATTTCAAAACCTTCACCGCCAC 1540
CATGTTTCAACTCTTAATTTTATAAAATTTTGCAATTTACAAATCGCCATCCCTTGCCTGAAAAGTGGC 1610
CACCAAAAATCAATTTCTCGGCATCATAATGACTTTTAAATGTATGTGAGAAAACACAGAAGAGGCATAC 1680
AAATGACAGGGACAGGTTGTCTCTCTCTCCCTCCTCTCCCGCTCCTCCTCCCGTCCATCTCCAAC 1750
1760 1770 1780 1790 1800 1810 1820
AACAACAATTTTCCAAATTCGTTGTCCATTTTGCCTATAAACAATTTGTGTGTGGAAAGGAAATACACGGG 1820
GAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTAATCTTTCCGAAATTGATCAATAAAGATAGAGC 1890
CGATGACACTGGCTGGTAGTAGTATGAGTGTAGAAATGCTTTTCATCGTCTCAACTTGCCTCATCAGTCT 1960
TCCCCGCTCTCATCACTGACAATTAATGTGGGGTTTTATGCGCTCTTTCCATTCGCGCACATCTCTG 2030
GGTTACCAACAACTGGAAATACATTTTACTACTATTCAAGCCATTATTTTGATATTTAAATTTGTGCAAT 2100
2110 2120 2130 2140 2150 2160 2170
TAGGGATAAACCAGACTTTTAAAGTTTATTTAAAAAACGATATTTTCGATTTTAAAAAATCTGAAAAG 2170
TTTCAAAAAATCAATAAATAATCCCAACAAATTTGATGGCTAAAAATTTTATTTCTACTGTGACAATAT 2240
CTTTATATGTATCACATGTTTTCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAGGCTCCGTTGCACA 2310
TTCCCATGCTATGAATGCTATGAGGACATATCCAAAAATTAAGCTAGACGGTTGATAATTAATGGGC 2380
ACGCGTAAATAAGTGCAAGCAGTTAGAATTTAAATCAAGCACAGATTATCTATCAAAATCAATCTTGA 2450

Friday, 28 November 1997 12:03
pBS KS/X16

Page

2460 2470 2480 2490 2500 2510 2520
ACATTCAGCCAGTTCGTACAATTTTCCATGCTTTTGGCCCATTAACAAACATTCCTCACCTCTTCAACCA 2520
TCTCACTCGTATCAATAAAGATATAGCAAAAGCCCGACTCTACTTTTTAAGAGAAGGAGATACTGAGCCA 2590
CAIGGCGTGAGACCTTTTCATCTCGTCCGTTCCGGTCTCAAAATTCACGCATATACTAACTCTTCAAAATAG 2650
CCATAGACCTCCCTTGTTTTCTTTCTTTGACTCGCGCCTATTTTTTGTTGGCTGGCTGAAAGCCGGGA 2720
AAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATAAAAAACGAGGCAATTTAAAAATAATTAATA 2800
2810 2820 2830 2840 2850 2860 2870
TTAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAACAAACAAATAGGAACCGCCAGATCAAA 2870
ATTCTATTTAAAGGTTTCAAGATGTTTAGGCAAGATTCCGGCTGAACAGAAAACCTGAAGTGCCATCATAA 2940
ATCTAGTGTAAAGTTTAGATTGAACCTCGAAATCTTAAGCCATGAATATAGCCTTATTCTAGATCTTAGT 3010
TGGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCATTTAGTTAAGCCATGATTGACTTGTCTGCTTC 3080
AGTCTAATCCAGACTAGATTTCCAAGAGAGTTTCAATTTAAATGTTTCCAGTTTCTTGTACTTAAAA 3150
3160 3170 3180 3190 3200 3210 3220
TCTTAAATGCGCTGTGATGCGTAAATCGTTATCCCCTTCTCTCACACTTTCAATTACAGATTCATCAAAAG 3220
ATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCACCTCATCAACCACTTCATCAAAATAATACAA 3290
ATTCATTCCGTCCGTTCGAGCCGTTCGAGTGGCAATAATAATGTTGGCTCGACGATATCCACATCTGCGAA 3360
GAGCTTAGGTATCCGATCCCTTCGCGCTTCTTTTTAGAAATATATATATTTTCAAGATCATCATCAACGAT 3430
AGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAACCTCTAGACCACAAACCCAGCTAG 3500
3510 3520 3530 3540 3550 3560 3570
TTCTGTGTTCTACAACTACAAAAATCGGAAGCTCAAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAA 3570
ACTTGCTTCTGTGAAGACTATTGGAGCAAAACAGAGCCCGATAACAGCGGTGGTGCTGGTGGTGAATG 3640
CTGAAATTAAGTTATTTCAGTAGCAAAAACCCATCTTCTCATCGAATAGCCCAACACCTACGAGAAAGG 3710
CGCCGGCGGTGCTCTAACAAACAACTTTTGTCGAAAAATCGCTGCCCCAGTGAAAAGTGGCTGAAAGCCGCC 3780
GACCAGTAAGCTGGGAAGTGGCAAGCTCTATGTCGAAGCTTTGTACGGTGAGTATTTTAAATCGGAAATG 3850
3860 3870 3880 3890 3900 3910 3920
GAAATGTATTTTAAAAAATGAAATTTCTACAAAATAAAATAAAAAATAGATTTTCTCTCTGATAGTA 3920
TTGCATCCCACTATTTTACATTTGAAGATTTATATCTTGGTTTCATATTGAAGATATCAGATAAGAAAG 3990
AAATAAAAAATATTTTGACAGTTGATAATTTTGTGATAGGACCAAGACAAAGTGAGATATAAGCTGTCAA 4060
AGTTGATTTTCAAGAAATTTTAAACCCCTAGTTTTCGAAGCTCTGGGCTCATCTATATTAGAACCCG 4130
ATTCGTACTTCTTCCGTTCCTTGACTCTACCAAAACCAAAACCAACCTACTAATAAATGATGAGACAA 4200
4210 4220 4230 4240 4250 4260 4270
TTGGGAATGTCTCCCAATTTCT 4270
GTCTGTCT 4340
ACGTCGAGGCGCCGAAAAACACATCTGGCTGACAAAGAGTAAAGCAATTTCTCAGCTTTCTCTCTCTCT 4410
TTTCAATTCGTTTTTCAAAATGAGCTACTACAGAGTGAAAGAGCACAAATTGCAAAACATTTTGTGTTGA 4480
GATGCACTTTTGA AAAAATTAACCTTACGTTTTCAGTTTCTAGTATTTATTTTTTATATAAATTAGAG 4550
4560 4570 4580 4590 4600 4610 4620
CTCTCTAGACCTGCTATATTTTAAAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4620
TGCTCTATGGCTCTCTATTTTATGAGAAAACATTTTTTAAAAATTTTTTGA AAAA AAAAACCTGCTCT 4690
TTCTTTTTTACATAGTAATTTCCAGCCAAAAGTTTCTTACCGTAAACGGACGCCCAATCATATCTCA 4760
ACAAGACTGAAGAGCAGTCAAGAGCAGTGAAGAAGAGTCCGGAATACGCTGGATTCACAGCAGCTG 4830
CCAACGTCTATCTGACGGAAGGTTCCCTAAGCATGCATTCACATCTTCCAAGGTCTGTTGTTTAGGAG 4900

Friday, 28 November 1997 12:03
pBS KS/X18

Page

4910 4920 4930 4940 4950 4960 4970
AACTGTTTTTTTTGTTTTGCTGACCTTCACATAGTCTCGGATGTTTATAAAAGTGAGGTCCTGGGACA 4970
CCTGCCATAAAATGTGAATCCGCCCAATTTGGTACAAAAAACTTTGCAGAGCACCTTGCTTTATACATT 5040
TTTAGGATAAAATGTCATACGGGATTTGTCAAAACCCAACTTTTTAAATTTTATTTTCAGATCAAAAATTG 5110
ATGTTAAAAGTTTAAAGATATTTACGAAAAAATGTTTACTTAAACCTTTTATATGGATAAAAATTTTAC 5180
AACATTCAGATAGGAGTTCCGTCCCTAAAACCTTTTGTGGTCGCCGAGAAGCTTTCGAATTAATAATGCA 5250
5260 5270 5280 5290 5300 5310 5320
TTTTTAAGTTCGAAACTAATTTTTTGTGCAACTGAATTTTTCAGAGATGAACCTTTAAGTTTCAATTTAT 5320
CCCAATTTGAACCCCTCCCTTCTATAAACTTCAAAATTTTCAGAGTTCAACGTCAGACGAAAAGTCTCC 5390
GTCTCAGACGATCTTACCTTAAACGCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCG 5460
GTCTCTCAAAATTTATCAACAAGCCTGTTGAGGTGAGTATTTTGTCTTCTGGGTCAGAGGCTCTTGTCT 5530
AAGTTTGGCCTAAATTAATTAACCTTGGTTTAGAGGCTGGCAAGCCATTGATCAAGCATGGGCTAATCT 5600
5610 5620 5630 5640 5650 5660 5670
GGGCCCGCCCTGAACCATGTACATATCTTTGGCCCGAGTAGTTGCAATCTAAAGATTCGAAGCTGGCTCT 5670
AAAGTCGGACTAGGCATAAAGTGCATAAAATGGAAAAATATCTTGAATTCATACGCTTCCGCTCTTCCAT 5740
CTCTCTCTTTTGTCTCTTTTGTTCAGATTTTCCCTTTTTTAGATTTTAAAGATTTTGATACACTTTTAA 5810
TGCTGGCTTCGCTTCTTAAGAGCCTTCTTATATTTTCGAAAAATAATCAATTTTAGGAAAAACCAAC 5880
ACTGGCAGTGAAGGAGTGAAGAGCACAGCGAAAAAAGATCCACCCTCCAGCTGCTTCGCCACGTGACACC 5950
5960 5970 5980 5990 6000 6010 6020
CAGCCAACAACTGGAGTTTGTAGTCCATTTATGGCACATAAGAAGTTGACAAATGGTACGTTTATTTCTG 6020
AACCTTACCTTATGTTTCTGGTCCGTGACGTTTTTGTTCACCATGTGATGGGAAGTAAATTTTGGATTATTT 6090
AAAGTTTGTGCGGGAAATAGTAAGGAGGAGTACAAATTTTATTTGTAGAAGGTTCTAGTAAACCTTTGAT 6160
TTTCTGACCATAAGTTTTTTTCTGAAAGTTGTTTAAAAAATTCAGTTAAAAAATAAAAIAATCTCTA 6230
TAAAAATTTCTAAATTTCTTGGGAATTTTTTCAAAAATGTTTTTCCAAAATAGTCTAATAGTAAGATTTG 6300
6310 6320 6330 6340 6350 6360 6370
TTTGTGAATTTACAAAAATATTTTAAAAATACATTTTAAATTTTATTCGATTTTTTCTGTTCCGACGAAACGAC 6370
AAAAAATCAGAAAAAGCGAAATTTAATTTCAAAAAAATATTTTGAAAAATCACAAATTAAGCTATTTTT 6440
CAAAAAATCAACAAAAAATATCAAAAGATTCTATTTTTCAGAAATAGAGACATAATCAAAAAATCAACA 6510
AAAAATTCATTTTTCCGGAAAAAATTCAGAAAAATTCAAAAATTTGAAAAAATCAAAATTCAGAAAAA 6580
ATTCAGAAATTTAAAAAATTTCTTTGAGGAAAAATTTAAAAATTTTAAATGTGTATTTCTGAAACCA 6650
6660 6670 6680 6690 6700 6710 6720
AGCATTTTCCGACTTTTTCGCGATTTTTCAGACTTGGGCTATAAATTTTTGTCAAAAATTAAGAAATCTTAAA 6720
ATATTTGTATTTTTTCGAAGAATCTCTCTCAATCTCAAAATTCATATTTTATAATTTTCAGACCCCGTGATAT 6790
CTGAAAAACAGAACCTGAAAAAGCTCCAAATCAATGAGCATTCGACACGAGGAGCTTTCACCGCTTCACAC 6860
TCTAAAAATCAGTTGTTTCCACTTAAAAATGACTTCAATCCGACAACCAACCTACGATGTTCTCTTAAAA 6930
CAAGGAAAAATCACATCGCTGTCAAGTCGTTTGGTCAAGTGCACCCCCCACTCCCAATTATATGACAAA 7000
7010 7020 7030 7040 7050 7060 7070
TGACCATTTTCAGGATATGAGCAGTCTGCGCGTCTGAAGACTCCATTGTGGCTCATGCGTCCGCTCAG 7070
GTGACTCCGCCGACAAAAACTTCTGGTAAATCATTCGCTTCGAGAGAAGGATGSGAAAATAAAGACAACAG 7140
GTAAATTTTGAAGAACCTTGATTTTTTTGTGAAAAATAGCTTCAAAATATAAATTTTAAAAAATCCGAG 7210
AAAAATGATGTTTGTCAAGGAAAACTTTGATTTTTTGGTTTCTGAACCTGTTTCGTTTAAAGTTAACCA 7280
ACGTTGGAGCTCGTACCAAAAACTTTTTCTTTGATAAATTTTGAATCTATAGTATTTTAAATTTTGGAA 7350

Friday, 28 November 1997 12:03
pBS KS/X16

Page

7360 7370 7380 7390 7400 7410 7420
TTGTGAAAGTTCCTTGAGATGTATTAAGTCTTAGGCATAGGCAGGTGTGTAGGCAAGAAAGGTATGATGT 7420
AGGCAGATAGGCTTGATATTTACCAAGCCAAATAAACAGTAAATAATATTTAAAAAAAGAACACTGAATAAA 7430
TCAAAAGCTAATAATATTTGTTTATTGGACCTACCAACACCTTACATTTGCCTACAAGCTTACCTATTTCC 7440
TTGTTGTCTACATTTGAACGTTAATCACTAATTCGGTGAATGAACACTTGTAGATTTTAATTTTCGACA 7450
GTAATTTTGGAGCATTGGCGTTAGAAATTTGAAAAAAACCTTCGACAGTTGAATCCTCATAACTCTCA 7460
7710 7720 7730 7740 7750 7760 7770
AAATAATTCAGAATCCAGCGGCTACACCTCTGACGCCGGTGTTCGATGTGCGCCAAATGAGGGAGAAG 7770
CTGAAAGAAATACGATGACATGACTCGTCGAGCAGAGAACCGCTATCCTGACAAGTGAATTTTGGTATGAT 7780
TAGTTGTAGTCCCTTGACACACATATGAACACATTCGCTGCTCGTTTCGGTGGTCAAGGAGCCATGGAGC 7790
AATTAATCCAGAAGGCTCAAAATTAATGAGCATCACTTGGTGAATCGAGGAATCCCCGAAAGACGTTTGGAT 7800
AGCATCTTCTTCCTTGCATTCTTTCTCTCTCTGCTGGGAGTCCCTGTTACACAGACATCTATTCTATG 7810
8080 8090 8100 8110 8120
CGCGAAGTGCAATTTTGGTTCCTAAAGATGAGAGGAGGAGGAGGAGCTATGACATATGAATGAGCATCGA 8120
GGAACCGCGTGAAATAGTTTGGAGCTTAGATTGCAAATTACAGGAATATTCGGTGAACCTCAGTCTACCT 8130
GGCATTTCGGCGGCGAGTTTGGTAAACATGAGCCAAATTTTGGTTCAGCGGTAATTTTCAAACAGTG 8140
GCAAGAAATTCCTCAACAAGTCTTGGCTGGCCCAAGTCTTCTAGAGTCTGATCCCAAGCTTGGTCCAATAC 8150
TTTTTGGGCCAAACTTTGGGAAGAAGTTGCCGCCAAATGCATTTTCAAATACTAGAACCAACATCT 8160
8410 8420 8430 8440 8450 8460 8470
GGTTTAGCGCAAGTTGTCTCCTAGGAGGATATCCAAATGTCTATTTACTCTCTCTCTTTCTAACGAAGAT 8470
CTATCGGTGCATCTCAATATGGAACCGATGCGGTTTTGCGCGCTTCGAAGAGCATTTCTTTGCTTTTGT 8480
GCCTTTTGTATGCTCTGAATTTGAACTAGTTATGAAATTCAGATGTTTCACTGGAGATGGCCAAAG 8490
ATAGTTGGTATTACGAAAGTTTTCATGAATTTTGAATTTCTGTAGTTCTGTGAGTCTCATACTATAGCTC 8500
AAAAAGTTCCGCTTGAGCCCTGCCATAATTCAAATTTCTTTCAAGTCCCTGCTTCTTAATTTAGGTT 8510
8760 8770 8780 8790 8800 8810 8820
ATGAAATAACGCGAACCATCATTTCTCTCTCTCAATTTCTTTTTCATGTCACCCTGATCCCTGCGGT 8820
TTCAATCTCTCTTCAATTCACCTTAGATTTGGCGCACAAGCTCCCTCTCTCTCTAATGCAAGCTGCTCT 8830
CGGCTTCCATTTTGGCGACTTGCCTCTTCCCTTCCCTTCCGCTCGCGGTTTCTTCTCGGATTGCCCTGT 8840
GTCTCTATAATTAATTCATTAAGCCGAGAACAAACGAACGGGGTTCTTTTCTCTCTCTTACCATCTTTT 8850
TGGCGGAATTAATTCGTCTTGTGATAGTGAAGTGTCTCAAGAATTTTGGTCTTGTGATAGCTTGGCA 8860
9110 9120 9130 9140 9150 9160 9170
AATATTGTTAGTATGCACCATGTTGCTCAGCATTTGTGTTCAAATCTTGGTTTATTTCAAATTTTGT 9170
TCAATAAACTTTGTAGGAGGTTTCTGTCGGCGGACCCACTTTTCAAATCTGAGCAAAAAATGCTG 9180
TCAATTCGATCTCAAAATATCTCAACTATAACCTTAGAACGTTTCTTCAAATTTCCGGAATCAAT 9190
TTTTCGAAATTTCAATTTTAAAGATTAACCTGCCAATTTGAATTCATGCGGCTTCCATTTGACTC 9200
CCATCATGACTACGTTCCGTEGGTCTCGCCACAAAAATACAGTTCCCTGGAAGGTTTCTTGTGGCGGGC 9210
9460 9470 9480 9490 9500 9510 9520
CTAGCAGAGCCCATAGGTCGCAAAAGCTCCCGATGAGCTGATAAAAAATGTACTCTGAGGAAATATATGC 9520
AAAAATTAATTTGAGGCTTTTACTAGAAACAGTTTAAGAAAGAAAAAGGTTTTTTTAAATTAATAA 9530
TTCAAATTTTGAATATTAAAGCCAAATTTGAGTTGATCACCTGAGAGAAATTCAAATATGAAAGGCTAG 9540
ATTTTGAATGCAATTTATCTGGAAGCGAATCTCTAAAATGAGAAAAACCTAGAAAAAATCTAAATGTTT 9550
GAAATTTTATTTGAAATCTAGTTGACTTTTTTGTATTTTCTAATTTGTTTCCAGGTAATAGTCTTT 9560

Friday, 28 November 1997 12:03
pBS KS/X16

Page

9810 9820 9830 9840 9850 9860 9870
GAAGTCGCTCCTCAACCTCAAACCAAGTGTGCTCCATATTTGGAACACACACAAACAAAAACCAATGA 9870
TAGTATGTGTTTCGAGTAGCCACTTGACAAGAAGAACTTGCCGACACIGGIGGCTGGTCACCATTCCTCT 9940
CTCTTTGTCAATTTGCAAAATCTTTCTCCCTCTTCTCTCAAAATAACTAACTGTGTGTCTGCGCGTCCC 10010
CGGCTCTCGAGGGGGGGGGCGGTACCCAGCTTTTGTTCCTTTTGTGAGGAGGTTAATTGCGCGCTTGGCGT 10080
AATCATGGTCATAGCTGTTTCTGTGTCAAATGTTATCCGCTCACAATTCACACAAACATACGAGCCCG 10150
10160 10170 10180 10190 10200 10210 10220
AAGCATAAAGTGTAAAGGCTGGGGTGGCTAATGAGTGAGCTAACTCACATTAATTGCGTTCGCGCTCAGTG 10220
CGCGCTTTCCAGTCGGGAACCTGTCTGTCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGGCG 10290
GTTTGGGTATTGGGCGCTCTTCGCTTCCTGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGGCG 10360
AGCGGTATCAGCTCACTCAAAGGCGGTAAACGGTTATCCACAGAAATCAGGGGATAACCCAGGAAAGAAC 10430
ATGTGACCAAAAGGCGAGCAAAAGGCCAGGAACCGTAAAAAGGCGCGTTGCTGGCGTTCCTTCATAGGC 10500
10510 10520 10530 10540 10550 10560 10570
TCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATA 10570
AAGAATCCAGSGGTTTCCCCCTGAAAGCTCCCTCGTGGCGTCTTCCGTTCCGAGCCGCGCGTTACCGGA 10640
TACCCTGTCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAGGCTGTAGGTATCTCAGTT 10710
CGGTGTAGGTCGTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGGCTT 10780
ATCCGCTAACTATCGCTTTGAGTCCAAACCGGTAAAGACAGGACTTATCGCCACTGGCAGCAGCCACTGGT 10850
10860 10870 10880 10890 10900 10910 10920
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGCCCTAACGAGGCT 10920
ACACTAGAAAGGACAGTATTTGGTATCTGGGCTCTGCTGAAGCCAGTTACCCTCGGAAAAGAGTTGGTAG 10990
CTCTTGATCCGGCAACAAACACCGCGTGGTAGCGGTGGTTTTTTTGTTTGAAGCAGCAGATTACCGCG 11060
AGAAAAAAGGATCTCAAGAAGATCTTTGATCTTTCTACCGGGTCTGACGCTCAATGGAACGAAAAC 11130
CAGGTTAAGSGATTTTGGTCAATGAGATATCAAAAAGSATCTTCACCTAGATCTTTTAAATTAATAATG 11200
11210 11220 11230 11240 11250 11260 11270
AAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATCTTAATCAGTGAG 11270
GCACCTATCTCAGCGATCTGCTATTTCTGTCATCCATAGTTGCTGACTCCCCGTCTGTAGATAACATA 11340
CGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCCAAGCTCAGCGGCTCC 11410
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGGCGAGCGCAGAAGTGGTCTGAACTTTATCCGCG 11480
TCCATCCAGTCTATTAATTGTTTCCCGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAAAGTTTGGCAAC 11550
11560 11570 11580 11590 11600 11610 11620
TTGTTGCCATTCGCTACAGGCATCGTGGTGTACCGCTCGCTGTTTGGTATGGCTTCAATCAGCTCCGGTTC 11620
GCAACGATCAAGCGAGTTACATGATCCCCATGTGTGCAAAAAAGCGSTTAGCTCTCTCGGTCCTCCG 11690
ATCGTTGTCAGAAGTAAGTTGGCCGCAAGTTATCACATCAATGTTATGGCAGCACTGCAAAATCTCTTA 11760
CTGTCATGCCATCCGTAAGAATCTTTCTGTGACTGGTGAAGTACTCAACCAAGTCAATCTGAGAATAGTC 11830
TATGCGGCGACCGAGTTGCTCTTGGCCGGGCTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA 11900
11910 11920 11930 11940 11950 11960 11970
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTGAGATCCA 11970
GTTGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCGCGCTTCTGGGTG 12040
AGCAAAAACAGGAAGGCAAAATGCGCGCAAAAAAGGGAATAAGGGCCACACGGAAATGTAATACTCATA 12110
CTCTCTCTTTTCAATATTAATGAAGCAATATCAGGCTTATTGTCATGAGCGGATACATATTTGAAT 12180
GATTTAGAAAAATAAACAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAT 12250

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence (Fig 361)
Enzymes: 100 of 146 enzymes (Filtered)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V 100

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTGGC
CCTCCAGCGACTCATCAGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCGGAACGGCTGTTAAGTACTTCTTAGACGAATCCCAATCCGCAAAACGC
G G R . V V R E O N L S Y N K A R L D R Q L H E E S A . G . A F C 200

CTGCTTCGCGATGTACGGCCAGATATACGGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGTCTATATGCGCAACTGTAATAAATGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
A A S R C T G O I Y A L T L I I D . L L I V I N Y G V I S S . P I Y 300

TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTGTGTTGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A W L T A Q R P P P I D V N N D V C S H S 400

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
TTGGGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A 500

CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCACTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTGATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGTGCATAATCAGT
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H 600

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTTGAATCACGGGGATTTCGAAGTCTCCACCCATTGACGTCAA
AGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCGCACCTATCGCCAACTGAGTGGCCCTAAAGGTTGAGAGGTGGGGTAAGTGCAGTT
R Y Y H G D A V L A V H O W A V I A V . L T G I S K S P P H . R Q 700

TGGGAGTTTGTGTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCGCCCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAG
ACCTTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATGTTGAGGCGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G 800

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGSETTATCGAAATTAATACGACTCACTATAGGGAGACCAAGCTGGCTAGC
CAGATATATTGCTCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCCGACCGATCS
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A 900

GTTTAAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGATGCTAGCATGACTGGTGGACAGCAAAATGGGTGCGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACTGTGCTTTACCCAGCCCTAGACATGCTG
F K L K L T M G G S H H H H H H G M A S M T G G Q Q M G R D L Y D 1000

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 2

GATGACGATAAGGTACCTAAGGATCCAGTGTGGTGAATTCGTCAGATATCGAATTCCTGCAGCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCC
CTACTGCTATTCCATGGATTCTTAGGTCACACCACCTTAAGACGCTATAGCTTAAGGACGTCGGGGACGAGAAGTCGGTCTACACCTGGGTCTCAGG

100

insert pLM1

ORF pLM1

D D D K V P K D P V W V N S A D I E F L Q P L L F S Q M L D P E S

AGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGA
TCTCTTCTCCTGTCACGCTTACAGGACCTAGAGGCCGCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCAGGGTCCACTGAGTGTGAGGGGACCT

200

insert pLM1

ORF pLM1

Q R K R T V Q N V L D L R Q N L E E T M S S L R G S Q V T H S S L E

GATGACCTGCTACGACAGCGATGATGCCAACCCACGACGCTGTCCAGCCTCTCCAACCGCTCGTCCCTCTGTCTATGGCGCTATGGCCAGTCCAGTCCC
CTACTGGACGATGCTGTGCTACTACGGTTGGGTGCTGTCACAGGTCGGAGAGTTGGCGAGCAGGGGAGACAGTACCGCGATACCGGTCAAGGTCAAGG

300

insert pLM1

ORF pLM1

M T C Y D S D D A N P R S V S S L S N R S S P L S W R Y G Q S S P

CGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCTGGTACATGCACGGCGAACGGGCCACTACTCCACAC
GCCGACGTCGGACCACTGCGCGGGAGACACCCACCTCGACGGCGAGCCTCCCTGCGGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGT

400

insert pLM1

ORF pLM1

R L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H

CCATGCCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACA
GGTACGGGTACGCGTCGGGTCTGTCGAGTCGGTATAGAGGGCGGACCTCGACCACTTAGGGACCTGAGCCTACTCCACCTGGAATTCAGCCGATGTA

500

insert pLM1

ORF pLM1

T M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y F

GAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCC
CTCGCTGTCACTGGAGTACCCGTCTGGTACTGCTCTACTACTGTAGTGATGGCCGACCTACTTTCTGTCGAGGTAGTCATCACTGAGTCGCTACGG

600

insert pLM1

ORF pLM1

S D S D L M G K T M T E D D D I T T G V D E S S S I S S G L S D A

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 3

TCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAAGTCAACAATAGTGCTAC
AGTCTGTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTTCATGAGGGTGACGAAGAGCGTCCTTGAGTTGTTATCAGCATG

insert pLM1

ORF pLM1

S D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L

GCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTGTAGTAATCAGAGGAGAAAGCCCTAAAAAAGTGGAGTACGACAGTGGTAG
CGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAATCACTTAGTCTCTCTTCGGGGATTTTTGACCTCATGCTGTCACCATC

insert pLM1

ORF pLM1

R T D S E K R S L A E S G L S V F S E S E E K A P K K L E Y D S G S

CCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTATCCAAGGGTGGAGAAGTGAAGAGCCCATCAGCGT
GGACTTCACCTTGACCCGTGAAGATTACCGCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCCACCTCTTGACTTTTCGGGTAGTCGGAC

insert pLM1

ORF pLM1

L K M E P G T S K W R R E R P E S C D D S S K G G E L K K P I S L

GGCCACCCTGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAATTCCTCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAG
CCGGTGGGACCAAGGGACTTCTTCCCGTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCGGGTCTACGGGAGTTTCAGCGTCCGTTG

insert pLM1

ORF pLM1

G H P G S L K K G K T P P V A V T S P I T H T A O S A L K V A G Y

CTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCTCTCTGATGCTGGTGGGACCGCTGAGTGAAGTAA
GACTCCCGTTTCGATGCTGTTCCTTCGAACGTCACTTCTTATGACCCGAGGTTGCGAGGAGGAGACTACGACCAAGCCCTGGCGGACTCACTACGATT

insert pLM1

ORF pLM1

P E G K A T D K G K L A V K N T G L Q R S S S D A G R D R L S D A F

GAGCCCCCTCGGGCATTGCTCGCCCC TCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTCTGCCACAGGCACAGCCACTGTGATGCAAACTGGT
CTTCGGGGGAGCCCGTAACGAGCGGGAGGTGAAGCCC TAGGAAGCCGATGTTCTTCGGAGGAGGACGGTGTCCGTGTCGGTGACAGTACGTTTGACCA

insert pLM1

ORF pLM1

K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGG
CCAAGTCGGTGAGAGTCGTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCGATCGAATCTACAAAGGTTGTCGCGTCTCG 2000

insert pLM1

ORF pLM1

G S A T L S K I O K S S G I P V K P V N G R K T S L D V S N S A E

CAGGATTCCTGGCTCCTGGAGCCCGTTC TAACATCCAGTACCGCAGCCTGCCCGGCCAGCCAAGTCAAGTTC TATGAGCGTGACCGCGGGCGGGGTGG
GTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCA TGGCGTCGGACGGGCGCGGTTCGTTCAAGATACTCGCACTGGCCGCCCGCCCAAC 2400

insert pLM1

ORF pLM1

P G F L A P G A R S N I O Y R S L P R P A K S S S M S V T G G R G G

ACCTCGCCCTGTGAGCAGCAGATTGACCCAGTCTCCTCAGCACCAAGCAGGGAGGCCCTTACGCCCTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGT
TGGAGCGGGACACTCGTCGTCGTAAC TGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGC GGAAGGTCTGACTTCTCGGATGGTTCCATCGGTCA 2500

insert pLM1

ORF pLM1

P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S

GGGCGGACCACTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGA
CCCGCCTGGTGAGGTGCGGGACAGTTAGTCTGTCTAGCCCTTTCTCTTCCGGTTTCGGTTCGCTCACCGGAACCTGAGTCTGTTGTAGAGGAAC TTCT 2600

insert pLM1

ORF pLM1

G R T T P A P V N Q T D R E K E K A K A K A V A L D S D N I S L I

GTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAA
CATAACCGAGGGGTCTCTCATGAGGGTTCTTGGTTCGTTTCGGTGGGGTGTGGTGGTTCGACCGTCTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCT 2700

insert pLM1

ORF pLM1

S I G S P E S T P K N Q A S H P T A T K L A E L P P T P L R A T A I

GAGCTTTGTCAAACACCCCTCAC TAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTG
CTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAAGTGTTCAGTTGAGGTGTGTCAGACCTAGATGGTAGGTACATATGGTGGGTACGAAGTTTCCAG 2800

insert pLM1

ORF pLM1

S F V K P P S L A N L D K V N S N S L D L P S S S D T T H A S I V

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 5

CCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCTGCTTCACCCCACTCCGGCACCCATCCTCAATATTAATCAGCCASCCTCT
GGCTAGACGTACGATGTTTCGAGTCGTAGACCCCGGGAGAGGGAAGGACGAAGTGGGGTCAGGCCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGA
3000
insert pLM1
ORF pLM1
P D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F
CCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGAT
GGTCCCGGACCTCGATTACTCACCAGGTCACACGGTTTCTCTGGGCGTACATGGGGTTGAGAGTCCGGACGTGTCTCGTACCTCAGGGAGGTC TA
3000
insert pLM1
ORF pLM1
S Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L Q M
GCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTGCTGCTCCACAGAAGAAGAGACGGAAGAGCTGACT
CGGTACTCGGAGGGGTACGGAAGGGTCTGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGCTTCTTCTCTGCTTCTCGACTGA
3100
insert pLM1
ORF pLM1
P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T
TGGAGTGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAAGGAGG
ACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCAATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTCCCGAGTCCATGGTCGAAGTCAGGGTCCTCC
3200
insert pLM1
ORF pLM1
S G S P R A G O L D S N Q R D R N T L P K K G L R Y Q L Q S Q E
AGACCAAGGAGAGGCGACATTCCCATACCATTTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACCTTCCCATGTCTCTGAG
CTGGTTCTCTCCGCTGTAAGGGTATGGTAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTC
3300
insert pLM1
ORF pLM1
E T K E R R H S H T I G G L P E S D D O S E L P S P P A L P M S L S
TGCAAGGGGCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCCAGCCCAAGAATCACCCGCTCCAAACAGCATCCCCACCCACGAGGGCGGCTTCGAG
ACSTTTCCCGGTGAATGGTTGTATCACTCAGGTGACGCCGGTGGTGGGTTCTTAGTGGGCGAGGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTC
3400
insert pLM1
ORF pLM1
A K G O L T N I V S P T A A T T P R I T R S N S I P T H E A A F E

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 6

CTGTACACGGCTCCCAATGGGGAGCACCCCTGTCCCTGGCCGAGAGACCCAGGGAATGATCGGTCAGGATCCTTCCGAGACCCACGGACGATGTT
GACATGTGCGCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCTTAGGAAGGCTCTGGGGTGCTTGTACAAAG 350

insert pLM1

ORF pLM1

L Y S G S Q M G S T L S L A E R P K G M I R S G S F R D P T D D V

ACGGCTCAGTGCTGTCCCTGGCTCCAGTGCTCCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAATCCGGAAGCTTCGTAGGGAAC
TGCCGAGTCACGACAGGGACCGGAGGTACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCTACGTTAGACTCGTTAGGCCCTCGAAGCATCCCTTGA 360

insert pLM1

ORF pLM1

H G S V L S L A S S A S S T Y S S A E E R M Q S E Q I R K L R R E L

GGAATCATCCAGGAAAAAGTGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCC
CCTTAGTAGGGTCTTTTTCACCGGTGGAAGTGCAGAGTCGAAAGACGGTTACGATTAGACCACGACGAAAACCTCGTCTCGGACCACTTATACGTAGG 370

insert pLM1

ORF pLM1

E S S Q E K V A T L T S Q L S A N A N L V A A F E Q S L V N M T S

CGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGTCTGGATTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCC
GCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCTGTGACTCGACGACCTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTCTTGAGACTCCGGG 380

insert pLM1

ORF pLM1

R L R H L A E T A E E K D T E L L D L R E T I D F L K K K N S E A

AGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACCTCTCAGATAGCATCTCAAGCCTCAACAG
TCCGTCASTAAGTCCCTCGGGAATACGGAGTCTTTGGTGTGGGTTCTTGAAGCCTAGTCTCTGTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTC 390

insert pLM1

ORF pLM1

Q A V I Q G A L N A S E T T P K E L R I K R Q N S S D S I S S L N S

CATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAA
GTAGTATCGGTAAGGTGCTAGCCGTCGTCGTTCTACGACTACGCTTTTCTCTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTT 400

insert pLM1

ORF pLM1

I T S H S S I G S S K D A D A K K K K K S V V Y E L R S S F N I

Tuesday, 18 November 1997 13:58

Page 7

pLM3 (1 > 10847) Site and Sequence

GGGTTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAAACTAC
CGCAAGTCATATTTTCCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGTTTGATG

insert pLM1

ORF pLM1

A F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L

AGCATGGTTCCACAGAGACTGCTTCACCTCCATCAAGTCTCCACCTTGCTCCGTGGGACATGATGTCACCGAGGGCCCTGCTACCCAGCCCCCA
TCGTACCAAGGTGTCTCTGACGAAGTGGGAGGTAGTTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGGT

insert pLM1

ORF pLM1

Q H G S T E T A S P S I K S S T L S S V G T D V T E G P A H P A P H

CACTAGGCTGTTCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGAC
GTGATCCGACAAGGTACGTTTACTCTCTCTCTCGGTCTTCTTCTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCTTTACTTCGAATGTCTG

insert pLM1

ORF pLM1

T R L F H A N E E E E P E K K E V S E L R S E L W E K E M K L T D

ATCCGCTTGGAGGCCCTCAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATG
TAGGCGAACCCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTCTTAC

insert pLM1

ORF pLM1

I R L E A L N S A H Q L D Q L R E T M H N M O L E V D L L K A E N

ACCGACTSAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCCTGGC
TGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCGTCACGGGACCTAGTAGACGTAAATAGAAGGGGTGCGGCGAGGGATCCGGACCC

insert pLM1

ORF pLM1

D R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S L G L A

ACTCACCCATTCTTCGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAGGAGGAAGTGACCTCCGGGTS
TGAGTGGGTAAGGAAGCCGGGGTCAGAACGCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCTCTTCACTGGGAGGCCAC

insert pLM1

ORF pLM1

L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GTGGTGAGGATGCCCCGACGACATCATCAAGGGGAC TTGAAGCAGCAGGAATCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGA
CAACACTCC TACGGGGGCGTCGTGTAGTAGTTCCCTGAACTTCGTCGTCCTTAAGAAGGACCCGACATCGTTCAGTCACCTTTTCACTGACCTTCT 4700

insert pLM1

ORF pLM1

V V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D V K

TGCTGGATGAAGCTGTTTCCAAGTGTTCAGGACTATATTTCTAAAATGGACCCAGCCTCTACCTGGGACTAAGCACTGAGTCCATCCATGGCTACAG
ACGACCTACTTCGACAAAAGGTTCAAGTTCCTGATATAAGATTTCCTGGGTCGGAGATGGGACCTGATTCTGACTCAGGTAGGTACCGATGTC 4800

insert pLM1

ORF pLM1

M L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

CATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCGAGATGCCCTCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGCTGAAG
GTAGTCGGTGCACTTTGCTCACAACCTACGTCCTGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTCCAGACTTC 4900

insert pLM1

ORF pLM1

I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L I

GAGAAATGCGTCGACAGCC TGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCTTGCTGAAGCACC GGCGCTCGTCCTCT
CTCTTTACGACAGTGTGCGACCAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTGCGAGGACGACTTCGTGGCCGCGGAGCAGGAGA 5000

insert pLM1

ORF pLM1

E K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L

CGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCA
GCCC GGGGTGGCGTGGCCGTTCGGATGGACTGGTTAGCGAACC GGCTCATGGACCACCTCGCGAGACC GGCACTCCAGTGCTCCCGTAGCAGTGGTG 5100

insert pLM1

ORF pLM1

S G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

CTTCAACATGCACAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTG
GAAGTTGTACGTGGTGGTCAGAACGTTCTTAGACGTTGACATAGAAAGGTTGATCGGTTGGTCTATCTGGCCCTTTGCTTAACCCCTACACGGGGAC 5200

insert pLM1

ORF pLM1

F N M H Q Q S C K D L O L Y L S N L A N Q I D R E T G I G D V P L

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACC TGCAAGTATCATAAATGTCCTTATATTATA3GTA
CACTAAGATAACCTACTGGACTCACCTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGTTCA TAGTATTACAGGGATATAATATCCAT 5300

----- insert pLM1 -----

----- ORF pLM1 -----

V I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G

CCACCAATCAGCCTGTAAAAATGACACCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCT
GGTGGTTAGTCGGACATTTTAC TGTGGGTTGGTACCGAACGTGAACCTCGAAGTCTACAAC TGGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGA 5400

----- insert pLM1 -----

----- ORF pLM1 -----

T T N Q P V K M T P N H G L H L S F R M L T F S N N V E P A N G F L

GGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGG
CCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTCTCGACGAAGCCACGAGCTGACCCATGGGTTCGACACC 5500

----- insert pLM1 -----

----- ORF pLM1 -----

V R Y L R R K L V E S D S D I N A N K E E L L R V L D V V P K L W

TATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCCATTGGCATTGAGGACTTCGGGA
ATAGTAGAGGTGTGGAAGGAACCTTCGTGTGCTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACTCCTGAAGSCCT 5600

----- insert pLM1 -----

----- ORF pLM1 -----

Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E D F R

CCTGGTTCATTGACCTGTGGAACAACCTATCATTCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGA
GGACCAAGTAAC TGGACACCTTGTGAGATAGTAAGGGATAGATGTCC TTCTCGGTTCTTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCTCCT 5700

----- insert pLM1 -----

----- ORF pLM1 -----

T W F I D L V N N S I I P Y L Q E G A K D G I K V H G Q K A A V E D

CCCAGTGGAATGGGTCCGGGACACACTTCCTTGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCCACCACCGTGGGCCCTCAGAGC
GGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTC TGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCTG 5800

----- insert pLM1 -----

----- ORF pLM1 -----

P V E W V R D T L P W P S A Q Q D Q S K L Y H L P P P T V G P H S

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 10

ATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTG
TAACGGAGTGGAGGGCTCCTATCCTGTGCTGCTGCTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACITTTGAAGTCTTCGAC 5900

insert pLM1

ORF pLM1

I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L Q E A

CCAAC TACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACGTGACCCCCGACAGCAGA
GGTTGATGTAAC TACAGAGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAAT TCCAAGCCGTTAGTGACAGTGGGGGCC TGTGCTC 3000

insert pLM1

ORF pLM1

A N Y I E S P D R E T I L D P N L Q A T L . G F G N H C H P R T A E

ACGCTGGCATCAGCTATCTTAGCTCCTCTCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAA
TGC GACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAGAAAGTCTCGTGACCGAGAGGTCGGGGTCTCTCTTGTCTCTCCCTCTCTACTT 6100

insert pLM1

R V H Q L S . L L L S P L L F O S T G S P A P G G E Q E G G G D E

AGAGGAGGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTTGTCGCCCTTAAACACATTTACTG
TCTCTCTCTCTGTCGAAGAACCACGACATGGAACCTCTTGAAGGATCCTCTTACCACCCACCGCAAACCTTGAACACGGGGGATTTGTGTAAATGAC 6200

insert pLM1

R G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T

GCCTCTCTAATGACTTTGGGGAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAACTCCTGGGCTTTCTGGGGAGGAGGTCAGAA
CGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAACCTGAAGAACAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCCAAGTCT 6300

insert pLM1

G L L . . L V G K D D S G S F P . L L V S I T N S W A F W G S V Q

AACATCAAAACACTGCAGCAGTTCCTCGGAATTCAGCTTGGACTTAACGAGCTGAACCTTGCTCAAAAGAAGCCGAATTCAGCAGACTGGCGGCCGTTA
TTGTAGTTTGTGACGTCGTCAAGGGGCCTTAAGTCGAACCTGAATTGGTCCGACTTGAAACGAGTTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAA 6400

insert pLM1

T S K H C S S S P E F S L D L T R L N L L K R S R I P A H W R P L

CTAGTTCTAGAGGGCCGTTTAAACCCGCTGATCAGCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCTCTTAC
GATCAAGATCTCCCGGGCAAATTTGGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAACGGGAGGGGGCACGGAAGGAAGT 6500

→

L V L E G P F K P A D Q P R L C L L V A S H L L F A P P P C L P .

CCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGAGGAAATTCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
GGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTACTCTTTAAGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCACCCACCC 6600

P W K V P L P L S F P N K M R K L H R I V . V G V I L F W G V G V G

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 11

CAGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTA
GTCTGTGCTTCCCCCTCCTAACCTTCTGTTATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCTTCTTGGTCGACCCCGAGAT 6700
R T A R G R I G K T I A G M L G M R V A L W L L R R K E P A G A L
GGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCCTTACACTTGCCAGCGCCCTAGCGCCCGC
CCCCCATAGGGGTGCGCGGACATCGCCGCGTAATTCGCGCGGCCACACCACCAATGCGCGTGCCTGCGGATGTGAACGGTTCGCGGGATCGCGGGCG 6800
G G I P T R P V A A H . A R R V W V L R A A . P L H L P A P . R P
TCCTTTCGCTTCTTCCCTTCTTCTCGCCACGTTGCGCGGCTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTAGTGCTTTA
AGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAAGCGGCCGAAAGGGGAGTTTCGAGATTAGCCCCGTAGGGAAATCCCAAGGCTAAATCACGAAAT 6900
L L S L S S L P F S P R S P A F P V K L . I G A S L . G S D L V L Y
CGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT
GCCGTGGAGCTGGGGTTTTTGAACATAATCCCACTACCAAGTGCATCACCGGTAGCGGGACTATCTGCCAAAAGCGGGAACTGCAACCTCAGGTGCA 7000
G T S T P K N L I R V M V H V V G H R P D R R F F A L . R W S P R
TCTTTAATAGTGGACTCTTGTTCAAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTGGGGATTTGGGCCTATTG
AGAAATTATCACCTGAGAACAAGGTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAATAAATATTCCTAAACCCCTAAAGCCGGATAAC 7100
S L I V D S C S K L E Q H S T L S R S I L L I Y K G F W G F R P I
GTTAAAAATGAGCTGATTTAACAAAAATTAACGCGAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGAAAGTCCCCAGGCTCCCCAGGCAGGC
CAATTTTTTACTCGACTAAATTGTTTTTAATTTGCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTCAGGGGTCCGAGGGTCCGTCGC 7200
G . K M S . F N K N L T R I N S V E C V S V R V W K V P R L P R Q A
AGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCAAGGTGTGAAAGTCCCCAGGCTCCCCAGCAGGCAGAGTATGCAAGCATGCATCTCAATT
TCTTCATACGTTTCTGACGTAGAGTTAATCAGTCTGTTGTCACACCTTTCAGGGGTCCGAGGGGTCTCCGCTTTCATACGTTTCTGACGTAGAGTTAA 7300
E V C K A C I S I S Q O P G V E S P Q A P Q Q A E V C K A C I S I
AGTCAGCAACCATAGTCCCGCCCTAAC TCCGCCATCCCGCCCTAACTCCGCCAGTTCCGCCATTCTCCGCCCATGGCTGACTAATTTTTTTTAT
TCAGTCGTTGGTATCAGGGCGGGATTGAGGCGGGTAGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCAGCTGATTAATAAAAAATA 7400
S Q O P . S R P . L R P S R P . L R P V P P I L R P M A D . F F L
TTATGACAGAGGCCGAGGCCCTCTGCC TCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTTAGGCTTTTGCAAAAAGCTCCCGGAGCT
AATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACTCTCCGAAAAAACCTCCGGATCCGAAAAAGTTTTTCGAGGGCCCTCGA 7500
F M Q R P R P P L P L S Y S R S S E E A F L E A . A F A K S S R E L
TGATATCCATTTTCGATCTGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCGCTTGGGTGG
ACATATAGGTAAAGCTAGACTAGTTCTCTGTCTACTCTAGCAAGCGTACTAATTTGTTCTACCTAACGTGCGTCCAAGAGGCGGGCAACCCACC 7600
V Y P F S D L I K R Q D E D R F A . L N K M D C T Q V L R P L G W
AGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCGTGTCCGGCTGTGACGCGAGGGGCGCCGGTCTTTTTTGCAA
TCTCCGATAAGCCGATACTGACCGTGTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTTCGCGTCCCCGCGGGCAAGAAAAACAGTT 7700
R G Y S A M T G H N R O S A A L M P P C S G C Q R R G A R F F L S
GACCGACCTGTCCGGTGCCTGAATGAAGTGCAGGACGAGGCGGCGGCTATCGTGGCTGGCCACGACGGCGTTCCTTGCGCAGCTGTGCTCGACGTT
CTGGCTGGACAGGCCACGGGACTTACTTGACGTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCGCAAGGAACGCGTTCGACACGAGCTGCAA 7800
R P T C P V P . M N C R T R Q R G Y R G V P R R A F L A O L C S T L

Tuesday, 18 November 1997 13:58
pLM3 (1> 10847) Site and Sequence

Page 12

GTCTAC TGAAGCGGGAAGGGAC TGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCGTCTTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACC 7800
S L K R E G T G C Y W A K C R G R I S C H L T L L L P R K Y P S W
CTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGC
GACTACGTTACGCCGCCGACGTATGCGAAC TAGGCCGATGGACGGGTAAGCTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCC TACC TTG 8000
L M Q C G G C I R L I R L P A H S T T K R N I A S S E H V L G W K
CGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACCTGTCGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGAT
GCCAGAACAGCTAGTCTACTAGACCTGCTTCTCGTAGTCCCGAGCGCGGTTCGGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGGTGCCGCTCTTA 8100
P V L S I R M I W T K S I R G S R O P N C S P G S R R A C P T A R I
CTCGTCGTGACCATGGCGATGCTGCTTGCCGAATATCATGGTGGAATATGGCCGCTTTTCTGGATTTCAGCTGTGGCCGGCTGGGTGTGGCGGACCC
GAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACC GGCGAAAAGACCTAAGTAGCTGACACCGCGCGACCCACACCGCTGG 8200
S S . P M A M P A C R I S W V K M A A F L D S S T V A G W V V R T
GCTATCAGGACATAGCGTTGGCTACCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGCTTTACGGTATCGCCGCTCCCGA
CGATAGTCTGTATCGAACCGATGGGCACTATAACGACTTCTCGAACCGCGCTTACCGGACTGGCGAAGGAGCAGAAATGCCATAGCGGCGAGGGCT 8300
A I R T . R W L P V I L L K S L A A N G L T A S S C F T V S P L P
TTCGACGCGCATCGCTTCTATCGCTTCTTGACGAGTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCAC
AAGCGTCGCGTAGCGGAAGATAGCGGAAGAAGCTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTG 8400
I R S A S P S I A F L T S S S E R D S G V R N D R P S D A Q P A I T
GAGATTTCGATTCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCT
CTCTAAAGCTAAGGTGGCGGCGGAAGATATTCCAACCCGAAGCCTTAGCAAAAGGCCCTCGCGCGACCTACTAGGAGGTGCGGCCCTAGAGTAGCA 8500
R F R F H R R L L . K V G L R N R F P G R R L D D P P A R G S H A
GGAGTTCTTCGCCCAACCCCACTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACG
CCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTAAAGTGTTTATTTCGTAATAAAGTAGC 8600
G V L R P P O L V Y C S L . W L Q I K O . H H K F H K . S ! F F T
CATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGT
GTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACC GCATTAGTACCAGTATCGACA 8700
A F . L W F V Q T H Q C I L S C L Y T V D L . L E L G V I M V I A V
TTCTGTGTGAAATGTTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCC TGGGGTGCTTAATGAGTGAGCTAAC TCAC
AAGGACACACTTAACAATAGCGAGGTGTAAGGTGTGTGTATGCTCGGCCTTCGATTTCACATTCGGACCCACGGATTACTACCTCGATTGAGTG 8800
S C V K L L S A H N S T Q H T S R K H K V . S L G C L M S E L T H
ATTAATTCGTTGCGTCACTGCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGGGGAGAGGGGTTTGCCT
TAATTAACGCAACGCGAGTGACGGGCGAAGGTCAGCCCTTTGGACAGCAGGTGACGTAATTACTTAGCCGGTTGCGCGCCCTCTCCGCCAAACGCA 8900
I N C V A L T A R F P V G K P V V P A A L M N R P T R G E R R F A
ATTGGGCGCTCTTCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTAT
TAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATA 9000
W A L F R F L A H . L A A L G R S A A A S G I S S L K G G N T V I

Tuesday, 10 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 13

CCACAGAATCAGGGGATAACGACGAAAGAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTTCCATAG
GGTGTCTTAGTCCCTTATTGCGTCTTTCTTGACACTCGTTTTCCGGTCGTTTTCCGGTCTTGCCATTTTTCCGGCGCAACGACCGCAAAAAGGTATC 910
H R I R G . R R K E H V S K R P A K G Q E P . K G R V A G V F P .
GCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCCCTGGAAGC
CGAGGCGGGGGGAC TGCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACGCTTTGGGCTGTCCTGATATTCTATGGTCCGCAAAAGGGGACCTTCG 920
A P P P . R A S Q K S T L K S E V A K P D R T I K I P G V S P V K
TCCCTCGTGCCTCTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTA
AGGGAGCACGCGAGAGGACAAGGCTGGGACGGGAATGGCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCACAT 930
L P R A L S C S D P A A Y R I P V R L S P F G K R G A F S M L T L
GGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAGCTGGGCTGTGTGCAGAACCCCGCTTCAGCCCGACCGCTGCGCTTATCCGGTAACATATCGTCT
CCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCCGACCGACACAGTGTCTGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGA 940
V S Q F G V G R S L Q A G L C A R T P R S A R P L R L I R . L S S
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
ACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCTGCGGTGACCATGTCTTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAAC 950
V Q P G K T R L I A T G S S H V . Q D . Q S E V C R R C Y R V L
AAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGAT
TTCACCACCGGATTGATGCCGATGTGATCTTCTGTCTATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATCGAGAACTA 960
E V V A . L R L H . K D S I V Y L R S A E A S Y L R K K S W . L L I
CCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
GGCGGTTTGTGGTGGCGACCATCGCCACCAAAAAACAAACGTTCTGCTGCTTAATGCGGCTCTTTTTTCTTAGAGTTCTTCTAGGAACTAGAAAAAG 970
R O T N H R W . R V F F C L Q A A D Y A Q K K R I S R R S F D L F
TACGGGCTGACGCTCAGTGAACGAAACTCACGTTAAGGGATTTTGGTCTAGAGATTACAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAA
ATGCCCCAGACTGCGAGTCACCTTGCTTTTGTGAGTGAATTCCTAAAACAGTACTCTAATAGTTTTCTAGAAGTGGATCTAGGAAATTTAATTTTT 980
Y G V . R S V E R K L T L R D F G H E I I K K D L H L D P F K L I
TGAAGTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT
ACTTCAAAATTTAGTTAGATTTCATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAG 990
M K F . I N L K Y I . V N L V . Q L P M L N Q . G T Y L S D L S I S
GTTCATCCATAGTTGCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACG
CAAGTAGGTATCAACGGACTGAGGGGACGACATCTATTGATGCTATGCCCTCCGAATGGTAGACCGGGGTACGACGTTACTATGGCGCTCTGGGTGC 1000
F I H S C L T P R R V D N Y D T G G L T I W P Q C C N D T A R P T
CTCACCGSCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAAT
GAGTGGCCGAGGTCTAAATAGTCTTATTTGGTGGTGGCTTCCCGGCTCGCGCTTCCACAGGACGTTGAAATAGGCGGAGGTAGGTACAGATAATTA 1010
L T G S R F I S N K P A S R K G R A Q K V S C N F I R L H P V Y
TGTGCGGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTA
ACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATATCAACGCGTTGCAACAACGGTAACGATGTCCTAGCACACAGTGGCAGCAGCAACCAT 1020
L L P G S . S K . F A S . F A Q R C C H C Y R H R G V T L V V W Y

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 14

TGGCTTCATTGAGCTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTAGCTCCTTCGGTCTCCGATCGTTGT
ACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCAATCGAGGAAGCCAGGAGGCTAGCAACA 1030
G F I Q L R F P T I K A S Y M I P H V V Q K S G . L L R S S D R C
CAGAAGTAAGTTGGCGCAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GTCTTCATTCAACCGCGCTACAATAGTGAGTACCAATACCGTCGTGACGTATTAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCA 1040
Q K . V G R S V I T H G Y G S T A . F S Y C H A I R K M L F C D W
GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGCGCTCAATACGGGATAATACCGCGCCACATAGCAGAAGT
CTCATGAGTTGGTTTCAGTAAGACTCTTATCACATACCGCGCTGGCTCAACGAGAACGGGCGCAGTTATGCCCTATTATGGCGCGGTGTATCGTCTTGAA 1050
V L N Q V I L R I V Y A A T E L L L P G V N T G . Y R A T . Q N F
TAAAAGTGCTCATCTATTGGAAAACGTTCTTCGGGGCGAAAACCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAA 1060
ATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACCTTAGGTCAAGCTACATTGGGTGAGCACGTGGGTT
K S A H H W K T F F G A K T L K D L T A V E I Q F D V T H S C T Q
CTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGT 1070
GACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAGACCCACTCGTTTTTGTCTTCCGTTTTACGGCGTTTTTCCCTTATTCCTCGCTGTGCTTTTACA
L I F S I F Y F H Q R F W V S K N R K A K C R K K G N K G D T E M
TGAATACCTACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAAC 1080
ACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCTATGTATAAACTTACATAAATCTTTTATTG
L N T H T L P F S I L L K H L S G L L S H E R I H I . M Y L E K . T
AAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGACGTC 10847
TTTATCCCAAGGCGCGTGTAAAGGGGCTTTTACGGTGGACTGCAG
N R G S A H I S P K S A T . R

FIG. 1.

Sequence of pTB72, an expression vector incorporating *C. elegans* UNC-53. The Open reading frame (ORF) of the prolinker + *Ce*-UNC-53 and (upper ORF) *Ce*-UNC-53 alone (lower ORF) are listed under the sequence.

modified from pTB72, do patent by changing ori on circle sequence; Created: maandag, 8 juli 1996 09:27:

5 AATGGCCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG
 10 TCATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA
 TTTCCAAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTGGCACCAAAATCAACGGGACTTCCAAAAATGT
 CGTAACAACTCCGCCCCATTGACGCCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTC
 TCTGGCTAACTAGAGAACCCACTGCTTACTGGGTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAAGCTT
 15 GGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTGCAGATATCCATCACACTGGCGG
 CCGCGGCCATGACGACGTCAATGTAGAAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTCGAAG
 GGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGGCACTATCGACTGGTTTCTCAGCTTATTAAT
 GTGATCGTTCGGATCAACGAATTTCTCGCTGCAATCGAAACGTTTGGCAAAATCACATCGAACCTGGATGG
 CCTCGAAACGTTGCTCGACTACCTGAAAAATCTGGGTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACA
 20 GCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTCTGCTCTCCACTACAAGCAGAAGCTTCGGCAACTGAAA
 AAAGATCAGAAGAAATGGAGCACTACCCACATCCATTATGCCACCGCGGTTTCTAAATACCTCGCCACGT
 GTCGCCACGTGACGAAACCGCTTCAGCAACTAACCCAAATTCGAATTTCCACAAATGTCAACATCCAGGCTTCAG
 ACTCCACAGTCAAGAAATATCGAAATTTGATTCATCAAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCA
 CGCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCGCTCGAGCCGTTTCGAGTGGCAATAATAAT
 GTTGGCTCGACGATATCCACATCTCGGAAGAGGTTAGAAATCATCATCAACGTACAGCTCTATTTCGAATCTAAAC
 25 CGACCTACTCCCAACTCCAAAAACCTTCTAGACCACAAACCCAGCTAGTTGCTGTTGCTACAACTACAAAAATC
 GGAAGCTCAAAGCTAGCGCGCTCCGAAAGCCGTGAGCACCCCAAACTTGCTTCTGTGAAGACTATTGGAGCAAA
 ACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGTGAATGCTGAATTAAGATTATTAGTAGCAAAAAACCCAT
 CTCTCTCATCGAATAGCCCAACACCTACGAGAAAGCGCGCGGCGGTGCTCAACAAACAACTTTGTCGAAATC
 GCTGCCCCAGTGAAGAGTGGCTGAAAGCGCGCCAGTAAAGCTGGGAAGTGGCCACGTCTATGTCGAAGCTTT
 30 GTACGCCAAAGTTTCTACCGTAAACCGGACGCCCAATCATATCTCAACAAGACTCGAAACGATGCTCAAAGA
 GCAGTGAAGAAAGTCCGGATACGCTGGATTCAACAGCACGTGCGCAACGTCATCATCGACGGAAGGTTCCCT
 AAGCATGCTATCCACATCTTCCAAAGAGTTCAACGTGACAGCGAAAGTCTCCGTCATCAGACGATCTTACTCTTAA
 CGCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCGT
 TTGAGGAAAAACCAACACTGGCAGTGAAAGGAGTGAAGACACGCGAAAAAGATCCACCTCCAGCTGTTCCG
 35 CCAGCTGACACCCAGCCAAACATCGGAGTTGTTAGTCCAATTATGGCACATAAGAGTTGACAAATGACCCCGT
 GATATCTGAAAAACCGAACCTGAAAAAGCTTCCCAATGAGCATCGACACGACGGACGTTCCACCGCTCCAC
 CTCTAAAAATCAGTTGTTCCACTTAAAAATGACTTCAATCCGACAAACCAACGATAGTGTCTTCTAAAAACAAGG
 AAAAAATCACATCGCTGTCAAGTCTGTTGATAGCAGCTGCTCCGCTCTGAAGACTCCATTGTGGCTCATG
 CGTCCGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGAAATAG
 40 ACATCAGAAATCCAGCGGTACACCTCTGACGCCGCTGTTGCGATGTGCGCCAAATGAGGGAGAAGCTGAAAG
 AATACGATGACATGACTCGTGCAGCACAGAACCGGTATCTGACAACTTCGAAGACAGTTCTCTTGTCTGCTG
 GGAATATCCGATAAACAAGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGGCAACAGTCCG
 TCCAAACATAGCGACTATCCCACTTTGTTGCCATCCACGCTCTTCTCTCAAGCCCGGATCCCGGCTCCG
 45 GGCTCCACATCAGTCGATTCTCGATCTCGAGCAGAAGAGGAGAAATGTGTACAACTTCTGTCCAGTCCGCGA
 ACGAGCCAAAGTGGCGCGCTGCCACCTCAACCTTCGGACAACATTGCTAAGATCCCGGGGATACTCATCCTA
 TTCTCCACACTATCAGTGTGAGCTGATAAGGACACAATGTCTATGCACTACAGACTAGTCCGACGACCTCTTC
 ACAAAAACCAAGCTATTACGGCAATTTCTATTCTGATCGTAAATGCCACCTTCAAGAGTTCAATCAGCTCCG
 50 GCACAGAAATGGCGGCTCTCTGAGCCCGAGACGGGTGCCGAACGATGTGCAAAATATGATTCTTCAGGATCC
 TACTCGCGCGGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGA
 TGAATAATCCCGGCACATTCTGCCAAAGTGAGATGGGATGCCAACTATCACTGGCTAGCAGCAGCATATGAT
 GATCTCTCAATGAGAAAGTACGAACATGCTATTCCGGGACATGGCAGTGACTTGGAGTGTACAAGAACATGTC
 55 GACTCACTAACCAAGAAACAGGAGAACTATGGACATTGTTGATCTTTTGAAGCAAAAGCTTAGAAAACTCACT
 CAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTACGGCAGGACATTGCTATTGAGGGA
 TATTAGCAATCATCTTGATCCAACTCAGCTCATGTCTAAGCAAGGCGCTGGTGAGCTTCTCTGTCACCATCTCT
 GGAATCAATTGTCATCCATCGATCATGATGTCTGTCGTAAGCAAGCAGCAAGCAGGAGAAGATCAGCTTGA
 GCTCGTTTGGCAAGAACAAAGAGCTGGATCCGCTCCTCACTCTCAAGTTCCAAAGAAAGAAAGAAACAAAG
 60 TACGACGAAGCAGATATGCCATCAATTTCCGATCTCAAGGAACCTTGTACAACTTGTGATGATGATGATGATG
 CAAGAGCTCAAGAAACGCGATAGTGCACTTTACGAAGTCCGCTTGCACATCTGGATCGTCCCGCGGAAGTTGA
 TGTCTGAGGGAGACAGTGAACAAAGTTGAAACCGGAGAACCAAGCAATTAAGAAAGAAAGTGGACAACTACCA
 ACGGTTCCAGCACTCGTCTCTTCCCGCGCTCAATTCAGTTATCTACGACGATGAGCATGTCTATGATGCA
 65 GCGTGTAGCAGTACATCAGCTAGTCAATCTTCAAAACGATCTCTGGCTGCAACTCAATCAAGGTTACTGTAAAC
 GTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTCCATGTC
 AACAGTCAAGTCATGCTGGAAGACATTGATGTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAAATGAT
 GTGGAGCATCAACTTGGAAATCGATGCTCGTGATCTTCTTGGCTATCAAAATGGTGAATTCGACGCGCTATT
 70 GGAGACTCCACAACTGATAACCAAGCCATCCAACTGACATTTCTACTCTCAACTACAAATCCGAATGTTATG
 CACGGTGGCGCACAGAGTCCGCTAGACAGTCTGGTCTTGGATATGCTTCTTCAAAAGCAAAATGATTCTCAACT
 CGTCAAGTCAATTTTACAGAGAGACGCTCTGGTGTAGCTGGAGCAACTGGAATTGGAAGAGCAAACTGGCGA
 AGACCTTGGCTGCTTATGATCTATTGCAACAAATCAATCCGAAGATAGTATTGTAATATCAGCATTCCTGAAAA
 AATAAAGAAAGAAATGCTTCAAGTGGAAACGACGCTGGAAAGAGTCTTGAAGCAAAAGAAATCATGCATCGTAAT
 TCTAGATAATATCCAAAGAAATCGAATTCGATTTGTTGATCCGTTTTTGCAAATGTCCCACTTCAAAACAGCA
 GGTCCATTTGTAGTATGCACAGTCAACCGATATCAAAATCCCTGAGCTTCAAAATCACCACAATTTGAAATGTCA
 GTAATTCGAATCGTCTCGAAGGATTCTCTACGTTACCTCCGACGAGCGGCGGTAGAGATGAGTATCGTCT
 AACTGTACAGATGCCATCAGAGCTTCAAAATCATTGACTTCTTCCAATAGCTCTTCCAGGCGTCAATAATTTT
 ATTGAGAAAAAGAAATCTGTTGATGTGACAGTTGGTCCAAGAGCATGCTTGAAGTGTCTCTCAACTGTCTGATGGA

FIG. 1 CONTINUED.

TCCCGTGAATGGTTCATTGCGATTGTGGAATGAGAACTTCATTCCATATTTGGAACGTGTGCTAGAGATGGCAAA
5 AAAACCTTCGGTTCGCTGCACTTCCTTCGAGGATCCACCGACATCGTCTCTAAAAAATGGCGGTGGTTCGATGG
TGAAAACCCGGAGAAATGTGCTCAAACGCTTCAACTCCAAGACCTCGTCCCGTCACTGCCAACTCATCCCGAC
AACACTCAATCCCCTCGAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATGCAACAACATTGAACAG
AAGACTCTAATCTTCTCTCGCCTCTCCCGCGTTTCCCTATCTTCTGACCGGTACCTGATGATTCCTCATTTCC
10 CCGTTTCCCGCCCAATTTCCGAGAACCTCCTGTTCCCTTTGTTCTAGTCTCTCCCGGTGCCGACGCCGAAGCG
ATTTAAAAACCTTTTCTTCCGAAACATTTCCTATGCTCAATTAAGTCAAATGAATAAACAGTGTATGTACTT
AAAAAAAAAAAAAAAAAACTCGAGGGGGGGGCTTCTATAGTGTACCTAAATGCTAGAGCTCGCTGATCAG
15 CCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCGGTCCCTTCTGACCTGGAAGGT
GCCACTCCACTGTCTTCTTAATAAAATGAGGAAATGCATCGCATTTGCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGTGGGCGAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGATGCTGGGGATGCGGT
GGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGTATCCCGCGCCCTGTAGCGG
CGCATTAAGCGCGGGGTGTGGTGGTTACGCGCAGCGTACCGCTACACTTCCGAGCGCTTAAATCGGGCATCC
20 TCCCTTCCGCTTCTTCCCTTCTTCTCGCCAGTTCGCGGCTTCCCGCTCAAGCTCTAAATCGGGCATCC
CTTAGGGTCCGATTAGTGTCTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGTGGTTCAGTAGT
GGGCCATCGCCCTGATAGACGGTTCCTGCGCTTTCGCGCTTTCGAGTGGAGTCCACGCTCTTAATAGTGAAGTCTT
25 CCAAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTGGGGATTTCGGCTA
TTGGTTAAAAATGAGCTGATTAAACAAAAATTAACGGCAATTAATCTGTGGAATGTGTGTCAGTTAGGTGT
GGAAAGTCCCAGGCTCCCAGGCGAGGAGATGCAAAAGCATGCATCTCAATTAGTCAGCAACAGGTTGT
GGAAAGTCCCAGGCTCCCAGGCGAGGAGATGCAAAAGCATGCATCTCAATTAGTCAGCAACAGGTTGT
30 GCGCTTAATCCGCGCATCCCGCCCTAACTCCGCGCATTCGCGCTTTCGCGCTTCTCCGCGCTTTCGCGCTA
TTTTTTATGAGAGGCGGAGGCGGCTCTGCTCTGAGCTATTCAGAGTAGTGAGGAGGCTTTTTTGA
GGCTAGGCTTTTGAAGAGCTCCCGGAGCTTGTATATCCATTTTCGATCTGATCAAGAGACAGTAGTAGG
ATCGTTTCGATGATTGAACAAGATGGATTGCACGCAAGTTCCTCGCGCTTGGGTGGAGAGGCTATTTCGG
25 TATGACTGGGCACAACAGACAATCGGCTGCTCTGATCGCGCGGTTCCTCGCGCTGTCAGCGCAGGGCGCGG
GTTCTTTTGTCAAGACCGACCTGTCCGCTGCTGATCGCGCGGTTCCTCGCGCTGTCAGCGCAGGGCGCGG
TGCCACGACGGCGGCTTCTTCCGCGAGCTGTGCTCGACGTTGCTACTGAAGCGGGAAGGAGTGGCTGCTATT
30 GGGCGAAGTGGCGGGCAGGATCTCTGTCATCTACCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
GCAATCGCGCGGCTGCATACGCTTGTATCGGCTACCTGCCATTTCGACCCAGCAAGCGAATCGCATCGAGC
GAGCAGTACTCGGATGGAAGCGGCTTGTGCTATCAGGATGATCTGAGCAGGAAGAGCATCAGGGGCTCGCGC
CAGCCGAATCTTCCGCGAGCTCAAGGCGCGCATCGCCGACGGCGAGGATGCTGCTGCTGCTGCTGCTGCT
35 CCGCTTTCGCGAATATCATGGTGGAAATGGCGCTTTCGCTGATTCTGCTGCTGCTGCTGCTGCTGCTGCT
GGACCGCTATCAGGACATAGCGTGGCTACCGCTGATATTGCTGAAGAGCTTGGCGCGCAATGGGCTGACCG
TTCTCTGCTTTACGGTATCGCGCTCCGATTCGCGAGCGCATCGCTTCTATCGCTTCTTACGAGTTCGATTCCA
40 CCGCGCTTCTATGAAAGGTTGGCTTCGGAATCGTTTCCGGGACGCGGCTGGAATGATCTCCAGCGCGG
GGATCTCATGCTGGAGTCTTCCGCCACCCCACTGTTTATTGAGCTTATAATGGTTACAAATAAGCAATAG
CATCACAATTCACAAATAAGCATTTTTTCACTGCAATCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTT
ATCATGCTGTATACCGTCACTCTAGCTAGAGCTTGGCGTAATCATGCTAGCTGTTTCTGTGTGAAAT
45 GTTATCGCTCACAATTCACACAACATACGAGCGGGAAGCATAAAGTGTAAAGCTGGGCTGCTAATGAGTG
AGCTAACTACATTAATTCGCTTCTGCTCACTGCGCGCTTTCAGTCCGGGAACCTGCTGCTGCTGCTGCTGCT
ATGAATCGGCAACGCGCGGAGGAGGCGGTTTCCAGTCCGGGAACCTGCTGCTGCTGCTGCTGCTGCTGCT
50 CTGCGCTCGGCTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAGGCGGTAATACGGTTATCCACAGAAT
CAGGGGATAACGCAAGGAAGCAATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTT
GCTGGCGTTTTTCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGAGCTCAAGTCAGAGTGGCGA
45 AACCCGACAGGACTATAAAGATACAGCGCTTCCCGCTGGAAGCTCCCTCGTGGCTCTCTGTTCCGACCT
GCGCTTACCGGATACCTGTCCGCTTCTCTCCCTCGGGAAGCGTGGCGCTTCTCAATGCTCAGCTGTAGG
TATCTCAGTTCGGTGTAGGTGCTTCCCTCAAGCTGGGCTGTGTGACGCAACCCCGCTTCCGCGCGCT
55 GCGCTTATCCGTAATCTGCTTGTAGTCCAACCGGTAAGACAGCACTTATCGCCACTGGCAGCAGCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCTAACTACGGCT
ACACTAGAAGGACAGTATTTGGTATCTGCGCTGCTGTAAGCCAGTTACCTCGGAAAGAGTGTGAGCTCT
TGATCCGGCAAAACAAACCGCTGGTAGCGGTGGTTTTTTGTTGCAAGCAGCAGATTACCGCGCAGAAAAA
60 AGGATCTCAAGAAGATCCTTTGATCTTTTACGGGGTCTGACGCTCAGTGGAAACGAAACTCAGTTAAGGA
TTTTGGTATGAGATTATCAAAAGGATCTTACCTAGATCTTTAAATTAATAAGTGTAAATCAATCTAA
AGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAAGCAGCTATCTCAGCGATCTGTCTA
55 TTTGCTTATCATAGTTGCTGACTCCCGCTGTAGATAACTACGATACGGGAGGCTTACATCTGGCC
CAGTGTGCAATGATACCGGAGAGCCAGCTACCGGCTCCAGATTATCAGCAATAAACCAGCCAGCGGAA
GGGCGAGCGCAGAAAGTGGTCTGCAACTTATCCGCTCCATCCAGTCTATTAATTTGTTGCGGGAAGCTAGA
60 GTAAGTAGTTCCGCAAGTAAATAGTTTGCACAAGCTTGTGCCATTGCTACAGGCTCGTGGTCAAGCTGCT
GTTTGGTATGGCTTCTTACGCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCGCTGTTGTGAAAA
AAGCGGTAGCTCCTTCCGCTCCGATCTGTCAGAAATGTAAGTTGGCGCAGTGTATCACTCATGGTTATG
65 GCAGCACTGCATAATCTCTTACTGTGATGCCATCGTAAGATGCTTTCTGTGACTGGTGAAGTCAACCAAG
TCATTCTGAGAATAGTATCGGCGACCGAGTGTCTTGGCCGGCTCAATACGGGATAATACCGCGCCACA
TAGCAGAACTTTAAAGTGTCTCATCTTGGAAACGTTCTTGGGGCGAAACTCTCAAGGATCTTACCGCTGT
70 GAGATCCAGTTCGATGTAACCCACTCGTGACCCCAACTGATCTTACGATCTTTTACTTTACCAAGCGTTTCTGG
GTGAGCAAAACAGGAAGGCAAAATCGCGCAAAAGGGAATAAGGGCGACAGCGAAATGTTGAATCTCATAC
TCTTCTTTTCAATATTTGAAGCATTTATCAGGTTTATGCTCATGAGCGGATACATATTTGAATGATTTAG
75 AAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCGAAAGTGCCACCTGACGCTGACGCGATCGGGAGATCT
CCCGATCCCTATGGTCACTCTCAGTACAATCTGCTGTGATGCCCATAGTTAAGCCAGTATCTGCTCCCTGC
TTGTGTTGGAGGTGCTGAGTAGTGCAGGCAAAATTAAGCTACAACAAGGCAAGGCTTACCGCAAT
GCATGAAGATCTGCTTAGGGTTAGGCGTTTTCGCTGCTTCCGATGTACGGGCGAGATACGCGTTGACAT
TGATTATTGACTAGTTATTAATAGTAATCAATACGGGCTCATTAGTTATAGCCCATATAGGAGTTCCGCTTA
CATACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTAT
75 GTTCCATAGTAACGCCAATAGGACTTCCATTGACGTCAATGGGTGGACTATTTACGGTAACGCGCACTT
GGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATGACGTCAATGACGGTA

FIG. 2.

Met Thr Thr Ser Asn Val Glu Leu Ile Pro Ile
Tyr Thr Asp Trp 15

Ala Asn Arg His Leu Ser Lys Gly Ser Leu Ser
Lys Ser Ile Arg 30

Asp Ile Ser Asn Asp Phe Arg Asp Tyr Arg Leu
Val Ser Gln Leu 45

Ile Asn Val Ile Val Pro Ile Asn Glu Phe Ser
Pro Ala Phe Thr 60

Lys Arg Leu Ala Lys Ile Thr Ser Asn Leu Asp
Gly Leu Glu Thr 75

Cys Leu Asp Tyr Leu Lys Asn Leu Gly Leu Asp
Cys Ser Lys Leu 90

Thr Lys Thr Asp Ile Asp Ser Gly Asn Leu Gly
Ala Val Leu Gln 105

Leu Leu Phe Leu Leu Ser Thr Tyr Lys Gln Lys
Leu Arg Gln Leu 120

Lys Lys Asp Gln Lys Lys Leu Glu Gln Leu Pro
Thr Ser Ile Met 135

Pro Pro Ala Val Ser Lys Leu Pro Ser Pro Arg
Val Ala Thr Ser 150

Ala Thr Ala Ser Ala Thr Asn Pro Asn Ser Asn
Phe Pro Gln Met 165

Ser Thr Ser Arg Leu Gln Thr Pro Gln Ser Arg
Ile Ser Lys Ile 180

Asp Ser Ser Lys Ile Gly Ile Lys Pro Lys Thr
Ser Gly Leu Lys 195

Pro Pro Ser Ser Ser Thr Thr Ser Ser Asn Asn
Thr Asn Ser Phe 210

Arg Pro Ser Ser Arg Ser Ser Gly Asn Asn Asn
Val Gly Ser Thr 225

Ile Ser Thr Ser Ala Lys Ser Leu Glu Ser Ser
Ser Thr Tyr Ser 240

FIG. 2 CONTINUED.

Ser Ile Ser Asn Leu Asn Arg Pro Thr Ser Gln
Leu Gln Lys Pro 255

Ser Arg Pro Gln Thr Gln Leu Val Arg Val Ala
Thr Thr Thr Lys 270

Ile Gly Ser Ser Lys Leu Ala Ala Pro Lys Ala
Val Ser Thr Pro 285

Lys Leu Ala Ser Val Lys Thr Ile Gly Ala Lys
Gln Glu Pro Asp 300

Asn Ser Gly Gly Gly Gly Gly Gly Met Leu Lys
Leu Lys Leu Phe 315

Ser Ser Lys Asn Pro Ser Ser Ser Ser Asn Ser
Pro Gln Pro Thr 330

Arg Lys Ala Ala Ala Val Pro Gln Gln Gln Thr
Leu Ser Lys Ile 345

Ala Ala Pro Val Lys Ser Gly Leu Lys Pro Pro
Thr Ser Lys Leu 360

Gly Ser Ala Thr Ser Met Ser Lys Leu Cys Thr
Pro Lys Val Ser 375

Tyr Arg Lys Thr Asp Ala Pro Ile Ile Ser Gln
Gln Asp Ser Lys 390

Arg Cys Ser Lys Ser Ser Glu Glu Glu Ser Gly
Tyr Ala Gly Phe 405

Asn Ser Thr Ser Pro Thr Ser Ser Ser Thr Glu
Gly Ser Leu Ser 420

Met His Ser Thr Ser Ser Lys Ser Ser Thr Ser
Asp Glu Lys Ser 435

Pro Ser Ser Asp Asp Leu Thr Leu Asn Ala Ser
Ile Val Thr Ala 450

Ile Arg Gln Pro Ile Ala Ala Thr Pro Val Ser
Pro Asn Ile Ile 465

Asn Lys Pro Val Glu Glu Lys Pro Thr Leu Ala
Val Lys Gly Val 480

Lys Ser Thr Ala Lys Lys Asp Pro Pro Pro Ala
Val Pro Pro Arg 495

Asp Thr Gln Pro Thr Ile Gly Val Val Ser Pro
Ile Met Ala His 510

FIG. 2 CONTINUED.

Lys Lys Leu Thr Asn Asp Pro Val Ile Ser Glu
Lys Pro Glu Pro 525

Glu Lys Leu Gln Ser Met Ser Ile Asp Thr Thr
Asp Val Pro Pro 540

Leu Pro Pro Leu Lys Ser Val Val Pro Leu Lys
Met Thr Ser Ile 555

Arg Gln Pro Pro Thr Tyr Asp Val Leu Leu Lys
Gln Gly Lys Ile 570

Thr Ser Pro Val Lys Ser Phe Gly Tyr Glu Gln
Ser Ser Ala Ser 585

Glu Asp Ser Ile Val Ala His Ala Ser Ala Gln
Val Thr Pro Pro 600

Thr Lys Thr Ser Gly Asn His Ser Leu Glu Arg
Arg Met Gly Lys 615

Asn Lys Thr Ser Glu Ser Ser Gly Tyr Thr Ser
Asp Ala Gly Val 630

Ala Met Cys Ala Lys Met Arg Glu Lys Leu Lys
Glu Tyr Asp Asp 645

Met Thr Arg Arg Ala Gln Asn Gly Tyr Pro Asp
Asn Phe Glu Asp 660

Ser Ser Ser Leu Ser Ser Gly Ile Ser Asp Asn
Asn Glu Leu Asp 675

Asp Ile Ser Thr Asp Asp Leu Ser Gly Val Asp
Met Ala Thr Val 690

Ala Ser Lys His Ser Asp Tyr Ser His Phe Val
Arg His Pro Thr 705

Ser Ser Ser Ser Lys Pro Arg Val Pro Ser Arg
Ser Ser Thr Ser 720

Val Asp Ser Arg Ser Arg Ala Glu Gln Glu Asn
Val Tyr Lys Leu 735

Leu Ser Gln Cys Arg Thr Ser Gln Arg Gly Ala
Ala Ala Thr Ser 750

Thr Phe Gly Gln His Ser Leu Arg Ser Pro Gly
Tyr Ser Ser Tyr 765

Ser Pro His Leu Ser Val Ser Ala Asp Lys Asp
Thr Met Ser Met 780

FIG. 2 CONTINUED.

His Ser Gln Thr Ser Arg Arg Pro Ser Ser Gln
Lys Pro Ser Tyr 795

Ser Gly Gln Phe His Ser Leu Asp Arg Lys Cys
His Leu Gln Glu 810

Phe Thr Ser Thr Glu His Arg Met Ala Ala Leu
Leu Ser Pro Arg 825

Arg Val Pro Asn Ser Met Ser Lys Tyr Asp Ser
Ser Gly Ser Tyr 840

Ser Ala Arg Ser Arg Gly Gly Ser Ser Thr Gly
Ile Tyr Gly Glu 855

Thr Phe Gln Leu His Arg Leu Ser Asp Glu Lys
Ser Pro Ala His 870

Ser Ala Lys Ser Glu Met Gly Ser Gln Leu Ser
Leu Ala Ser Thr 885

Thr Ala Tyr Gly Ser Leu Asn Glu Lys Tyr Glu
His Ala Ile Arg 900

Asp Met Ala Arg Asp Leu Glu Cys Tyr Lys Asn
Thr Val Asp Ser 915

Leu Thr Lys Lys Gln Glu Asn Tyr Gly Ala Leu
Phe Asp Leu Phe 930

Glu Gln Lys Leu Arg Lys Leu Thr Gln His Ile
Asp Arg Ser Asn 945

Leu Lys Pro Glu Glu Ala Ile Arg Phe Arg Gln
Asp Ile Ala His 960

Leu Arg Asp Ile Ser Asn His Leu Ala Ser Asn
Ser Ala His Ala 975

Asn Glu Gly Ala Gly Glu Leu Leu Arg Gln Pro
Ser Leu Glu Ser 990

Val Ala Ser His Arg Ser Ser Met Ser Ser Ser
Ser Lys Ser Ser 1005

Lys Gln Glu Lys Ile Ser Leu Ser Ser Phe Gly
Lys Asn Lys Lys 1020

Ser Trp Ile Arg Ser Ser Leu Ser Lys Phe Thr
Lys Lys Lys Asn 1035

Lys Asn Tyr Asp Glu Ala His Met Pro Ser Ile
Ser Gly Ser Gln 1050

FIG. 2 CONTINUED.

Gly Thr Leu Asp Asn Ile Asp Val Ile Glu Leu
Lys Gln Glu Leu 1065

Lys Glu Arg Asp Ser Ala Leu Tyr Glu Val Arg
Leu Asp Asn Leu 1080

Asp Arg Ala Arg Glu Val Asp Val Leu Arg Glu
Thr Val Asn Lys 1095

Leu Lys Thr Glu Asn Lys Gln Leu Lys Lys Glu
Val Asp Lys Leu 1110

Thr Asn Gly Pro Ala Thr Arg Ala Ser Ser Arg
Ala Ser Ile Pro 1125

Val Ile Tyr Asp Asp Glu His Val Tyr Asp Ala
Ala Cys Ser Ser 1140

Thr Ser Ala Ser Gln Ser Ser Lys Arg Ser Ser
Gly Cys Asn Ser 1155

Ile Lys Val Thr Val Asn Val Asp Ile Ala Gly
Glu Ile Ser Ser 1170

Ile Val Asn Pro Asp Lys Glu Ile Ile Val Gly
Tyr Leu Ala Met 1185

Ser Thr Ser Gln Ser Cys Trp Lys Asp Ile Asp
Val Ser Ile Leu 1200

Gly Leu Phe Glu Val Tyr Leu Ser Arg Ile Asp
Val Glu His Gln 1215

Leu Gly Ile Asp Ala Arg Asp Ser Ile Leu Gly
Tyr Gln Ile Gly 1230

Glu Leu Arg Arg Val Ile Gly Asp Ser Thr Thr
Met Ile Thr Ser 1245

His Pro Thr Asp Ile Leu Thr Ser Ser Thr Thr
Ile Arg Met Phe 1260

Met His Gly Ala Ala Gln Ser Arg Val Asp Ser
Leu Val Leu Asp 1275

Met Leu Leu Pro Lys Gln Met Ile Leu Gln Leu
Val Lys Ser Ile 1290

Leu Thr Glu Arg Arg Leu Val Leu Ala Gly Ala
Thr Gly Ile Gly 1305

Lys Ser Lys Leu Ala Lys Thr Leu Ala Ala Tyr
Val Ser Ile Arg 1320

FIG. 2 CONTINUED.

Thr Asn Gln Ser Glu Asp Ser Ile Val Asn Ile
Ser Ile Pro Glu 1335

Asn Asn Lys Glu Glu Leu Leu Gln Val Glu Arg
Arg Leu Glu Lys 1350

Ile Leu Arg Ser Lys Glu Ser Cys Ile Val Ile
Leu Asp Asn Ile 1365

Pro Lys Asn Arg Ile Ala Phe Val Val Ser Val
Phe Ala Asn Val 1380

Pro Leu Gln Asn Asn Glu Gly Pro Phe Val Val
Cys Thr Val Asn 1395

Arg Tyr Gln Ile Pro Glu Leu Gln Ile His His
Asn Phe Lys Met 1410

Ser Val Met Ser Asn Arg Leu Glu Gly Phe Ile
Leu Arg Tyr Leu 1425

Arg Arg Arg Ala Val Glu Asp Glu Tyr Arg Leu
Thr Val Gln Met 1440

Pro Ser Glu Leu Phe Lys Ile Ile Asp Phe Phe
Pro Ile Ala Leu 1455

Gln Ala Val Asn Asn Phe Ile Glu Lys Thr Asn
Ser Val Asp Val 1470

Thr Val Gly Pro Arg Ala Cys Leu Asn Cys Pro
Leu Thr Val Asp 1485

Gly Ser Arg Glu Trp Phe Ile Arg Leu Trp Asn
Glu Asn Phe Ile 1500

Pro Tyr Leu Glu Arg Val Ala Arg Asp Gly Lys
Lys Thr Phe Gly 1515

Arg Cys Thr Ser Phe Glu Asp Pro Thr Asp Ile
Val Ser Lys Lys 1530

Trp Pro Trp Phe Asp Gly Glu Asn Pro Glu Asn
Val Leu Lys Arg 1545

Leu Gln Leu Gln Asp Leu Val Pro Ser Pro Ala
Asn Ser Ser Arg 1560

Gln His Phe Asn Pro Leu Glu Ser Leu Ile Gln
Leu His Ala Thr 1575

Lys His Gln Thr Ile Asp Asn Ile

FIG. 3. : tblastn search of the EST division of Genbank with the ORF of the longest known *Ce*-UNC-53 cDNA, tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in *Ce*-UNC-53.

5 TBLASTN 1.4.9MP [26-March-1996] [Build 14:27:13 Apr 1 1996]

10 Reference: Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-10.

Query= tb3 M5 ORF
(1583 letters)

15 Database: Non-redundant Database of GenBank EST Division
647,253 sequences: 234,216,808 total letters.

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum P(N)	High Probability N
dbj10357801CELK025D6F C.elegans cDNA clone yk25d6 : 5'... +2	358	7.9e-54	3	
dbj10330481CELK025D6R C.elegans cDNA clone yk25d6 : 3'... -1	177	8.6e-16	1	
gb1H090361H09036 y196c11.r1 Homo sapiens cDNA clo... +1	115	1.1e-05	1	
gb1AA0491241AA049124 mj46f04.r1 Soares mouse embryo N... +3	106	8.6e-05	1	
gb1R914751R91475 yq08c11.r1 Homo sapiens cDNA clo... +2	59	0.21	2	
gb1T234461T23446 seq2955 Homo sapiens cDNA clone ... -1	61	0.99	2	
25 gb1R863901R86390 SWJICA339SK Brugia malayi infect... +2	74	0.996	1	
gb1T447811T44781 8044 Arabidopsis thaliana cDNA c... +1	71	0.9992	1	
gb1T755821T75582 yd63f11.r1 Homo sapiens cDNA clo... +2	64	0.99992	2	

30 gb1H090361H09036 y196c11.r1 Homo sapiens cDNA clone 46037 5'.
Length = 489

Plus Strand HSPs:

Score = 115 (52.1 bits), Expect = 1.1e-05, P = 1.1e-05

Identities = 22/70 (31%), Positives = 45/70 (64%), Frame = +1

35 Query: 1059 IELKQELKERDSALYEVRLDNLDRAREVDVLRITVNLKLTENKQKKVEVDKLTNGPATRA 1118
+L+ EL+++ L ++RL+ L A ++D LRE +N++++E ++LK D+L +
Sbjct: 7 MQLRNELRDKEMKLTDIRLEALSSAHOLDOLREAHNRMOSEIEKLVXNDRLKSESQSG 186

40 Query: 1119 SSRASIPVIY 1128
SR S P ++

Sbjct: 187 CSRGSFPSVH 216

45 gb1AA0491241AA049124 mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus
musculus cDNA clone 479167 5'
Length = 337

Plus Strand HSPs:

Score = 106 (48.0 bits), Expect = 8.6e-05, P = 8.6e-05

50 Identities = 23/58 (39%), Positives = 38/58 (65%), Frame = -3

Query: 1057 DVIELKQELKERDSALYEVRLDNLDRAREVDVLRITVNLKLTENKQKKVEVDKLTNGP 1114
+V EL+ EL E++ L ++RL+ L A ++D LRET++ ++ E LK E D+L P
55 Sbjct: 99 EVSELRLSELWEKEMKLTDIRLEALSSAHOLDOLRETMHNMHOLEVDLLKAENDRLKVP 272

FIG. 4.

A Search of the Genbank databases with part of the nucleotide binding domain of *Ce*-UNC-53 does not identify statistically significant proteins except for the *C. elegans* cosmid containing *Ce*-unc-53.

```

5  TBLASTN 1.4.0MP [20-June-1995] [Build 18:00:05 Aug 29 1995]
   Query= section5 (240 letters)
   >|cl|section5
   ILTERRLVLAGATGIGKSKLAKTLAAYVSIRTNQSEDSIVNISIPENNKEELLOVERRLE
10  KILRSKESCIVILDNIPKNRIAFVVSFANVPLONNEGPFVCTVNRVQIPELOIHNNFK
   MSVMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIEKTSVD
   VTVGPRACLNCPLTVGDSREWFIRLWNNFIPYLERVARDGKKNLRLSLHFLRGSHRRLX

   Database: Non-redundant PDB+CBupdate+GenBank+EMBLupdate+EMBL
             520,383 sequences; 367,017,413 total letters.

15                                     Smallest
                                     Sum
                                     Reading High Probability
                                     Frame Score P(N) N
   Sequences producing High-scoring Segment Pairs:
   embi247810ICEF45E1C Caenorhabditis elegans cosmid F45... -2 1131 5.1e-158 2
20  gbiR41071IR41071 Hk575-f Homo sapiens cDNA clone k... +2 53 0.33 2
   gbiT44781IT44781 8044 Arabidopsis thaliana cDNA cl... +1 74 0.35 1
   embi248334ICEF10B5 Caenorhabditis elegans cosmid F10... +3 71 0.83 3
   gbiM81884IEPFCPCG Epifagus virginiana chloroplast c... +1 49 0.91 4
   gbiL09547IPEAPCLP Pisum sativum (clone pCLP) nuclea... +1 71 0.99 1
25  gbiM32604ITOMCD4B Tomato ATP-dependent protease (CD... +1 71 0.99 1
   embiX69188IAPTUSGA A.pyhilitidis mRNA for gamma-tubulin +2 56 0.992 3
   gbiT44782IT44782 8045 Arabidopsis thaliana cDNA cl... +1 68 0.9995 1
   gbiM17087IHUMRASK12 Human c-ras-Ki-2 activated oncogene... +2 58 0.9998 1
   embiX57702IGSNATRIUP G.gallus RNA for precursor of nat... +3 56 0.9999 2
30  gbiK01520IHUMRASKB1 Human lung adenocarcinoma (PR371)... +2 57 0.99995 1

   >gbiR41071IR41071: Hk575-f Homo sapiens cDNA clone k575-f.
   Length = 310
   Plus Strand HSPs:

35  Score = 53 (24.5 bits), Expect = 0.40, Sum P(2) = 0.33
   Identities = 9/15 (60%), Positives = 13/15 (86%), Frame = +2

   Query: 130 GFILRYLRRRAVEDE 144
           GF++RYLRR+ VE +
40  Sbjct: 26 GFLVRYLRRKLVED 70

   Score = 47 (21.7 bits), Expect = 0.40, Sum P(2) = 0.33
   Identities = 9/26 (34%), Positives = 17/26 (65%), Frame = +3

45  Query: 170 NNFIEXTNSVDVTVGPRACLNCPLTV 195
           - F-EK +-+D -GP L+ PL +
   Sbjct: 147 HTFLEXHSTLDFLIGPCFFLSGPLAL 224

```

FIG. 5.

Three frame translation of EST gb:R41071.

Regions of homology region with *Ce-Unc-53* in two different frames are underlined. The spacing between the blocks of homology is of similar size as that in *Ce-UNC-53*.

- 5 Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with *Ce-UNC-53* (see alignment in fig 12).

10
15
20
25
30
35
40
45

```
CTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGT
  10      20      30      40      50      60
  L Q Q R G A S Q W L P G S L P E E E A G
  S N N V E P A N G F L V R Y L R R K L V
  P T T W S Q P M A S W F V T * G G S W *

AGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGGTGCTCGACTTGG
  70      80      90      100     110     120
  R V R Q R H Q C Q Q G R A A S G C S T W
  E S D S D I N A N K E E L L R G A R L G
  S Q T A T S M P T R K S C F G V L D L G

GTACCCAAGCCTGTGGTATCATCTTCCACACCTTCCTTGAGAAGCACAGCACCTTAGACT
  130     140     150     160     170     180
  V P K P V V S S S T P S L R S T A P * T
  Y P S L W Y H L P H L P * E A Q H L R L
  T Q A C G I I F H T F L E K H S T L D F

TTCTCATCGGCCCTTGCTTCTTTCTGTGCGGTCCATTGGCATTGAGGCTTCCGGACCTTG
  190     200     210     220     230     240
  F S S A L A S F C R V H W H * G F R T L
  S H R P L L L S V G S I G I E A S G P C
  L I G P C F F L S G P L A L R L P D L V

TTTATTGACCTGTGGACAACCTCTATCATTTCTTCTACAGGAGGAGCCAAGGATTGGAT
  250     260     270     280     290     300
  F I D L W T T L S F P I Y R R S Q G L D
  L L T C G Q L Y H F L S T G G A K D W I
  Y * P V D N S I I S Y L Q E E P R I G *

AAAGGTCCAT
  310
  K G P
  K V H
  R S
```

FIG. 6 : blastn search of the EST division of Genbank with Hu-unc-53/1 cDNA 3b.

5

10

15

20

25

30

35

```

BLASTN 1.4.9MP (26-March-1996) (Build 14:27:07 Apr 1 1996)
Query= Hu-unc-53/1 cDNA 3b
      (3256 letters)
Database= Non-redundant Database of GenBank EST Division
          647,253 sequences; 234,226,808 total letters.

```

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum P(N)	Probability	LOCUS N assignment
gb N36659 N36659 yx91b09.r1 Homo sapiens cDNA clone 2...	1668	2.1e-130	1	hu-UNC-
gb AA043997 AA043997 zk58a01.r1 Soares pregnant uterus Nb...	1316	8.3e-129	3	hu-UNC-53/1
gb AA049124 AA049124 mj46204.r1 Soares mouse embryo NbME1...	1324	9.1e-102	1	mm-UNC-53/1
gb T055601 T055601 ESTC3449 Homo sapiens cDNA clone HFB...	892	5.1e-84	3	hu-UNC-53/1
gb N24681 N24681 yx91b09.s1 Homo sapiens cDNA clone 2...	782	9.9e-75	2	hu-UNC-53/1
gb R41071 R41071 Hk575-f Homo sapiens cDNA clone k575-f...	535	1.5e-72	4	hu-UNC-53/1
gb N89104 N89104 K7846F Fetal heart, Lambda 2AP Expre...	451	7.3e-57	2	hu-UNC-53/1
gb R41073 R41073 Hk144-f Homo sapiens cDNA clone k144-f...	555	1.5e-36	1	hu-UNC-53/1
gb R15492 R15492 HH434-F Homo sapiens cDNA clone H434-F...	416	2.3e-29	2	hu-UNC-53/1
gb H09036 H09036 y196c11.r1 Homo sapiens cDNA clone 4...	438	9.4e-26	1	hu-UNC-53/2
gb W91567 W91567 MTA.C36.093.A MTA adult mouse thymus...	317	1.9e-17	2	mm-UNC-53/?
gb W74400 W74400 zd62a10.r1 Soares fetal heart NbHH19...	243	2.2e-09	1	hu-UNC-53/1
gb AA003314 AA003314 mg56h10.r1 Soares mouse embryo NbME1...	141	0.54	1	

FIG. 7.

TBLASTN search of the Genbank sequence database with the 961 aminoacid ORF of cDNA 3b of hu-UNC-53/1. hu-UNC-53/1 forms a unique pair with *Ce*-UNC-53 (cosmid F45E10) compared to the rest of the database.

5

10

TBLASTN 1.4.3MP (26-March-1996) (Build 14:27:13 Apr 1 1996)

15 Query= tmpseq_1
 (961 letters)

Database: Non-redundant GenBank+EMBL+DDBJ+PDB sequences
261,674 sequences; 371,416,172 total letters.

20

	Sequences producing	High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Probability P(N)	
25	emb1247810 CEF45E10	Caenorhabditis elegans cosmid F45E10	-2	158	2.5e-32	7
	gb1M97501 HUMCLIP	Human cytoplasmic linker protein...	+3	83	0.47	1
	emb1X64838 HSRESTIN	H.sapiens mRNA for restin	+1	83	0.47	1
	gb1M58752 ECONCRBC	E.coli mcrB and mcrC genes, compl...	+3	82	0.56	1
30	emb1Z11582 SCNUF1G	S.cerevisiae nuf1 gene	+1	82	0.61	2
	emb1X73297 SCSETRP4	S.cerevisiae spacer element	-1	82	0.74	2
	emb1X54002 XLKINESIN	X.laavis mRNA for kinesin	+2	63	0.85	5
	gb1U42409 DDU42409	Dictyostelium discoideum myosin h...	+3	66	0.92	2
	gb1U10399 YSC8082	Saccharomyces cerevisiae chromoso...	+2	77	0.93	2
	gb1U20810 ATU20810	Arabidopsis thaliana cytoskeleton...	+1	77	0.95	2
35	gb1L07879 LEIXINLIKE	Leishmania chagasi kinesin-like p...	+2	78	0.95	1
	gb1L03188 YSCINTANA	Saccharomyces cerevisiae integrin...	+2	65	0.997	2
	gb1U28372 YSCD9476	Saccharomyces cerevisiae chromoso...	+3	82	0.9991	2
	gb1M94362 HUMLAHBBA	Human lamin B2 (LAHB2) mRNA, part...	+1	75	0.9996	1
40	gb1M58337 VACHAGHA	Vaccinia virus hemagglutinin gene...	+1	74	0.99995	1

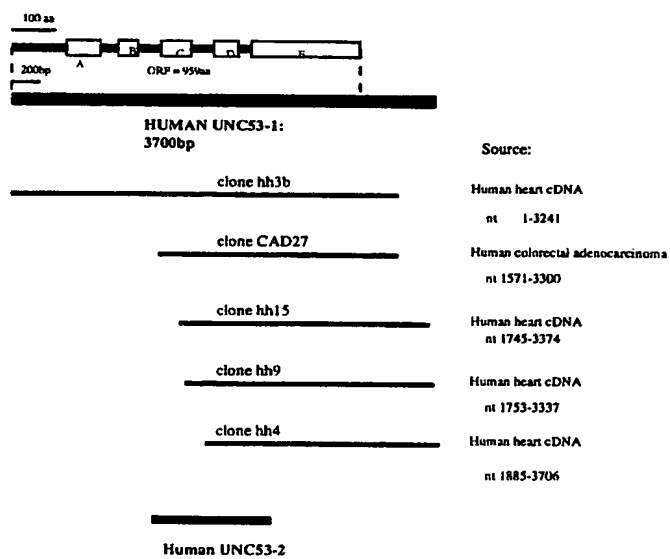


Figure 8

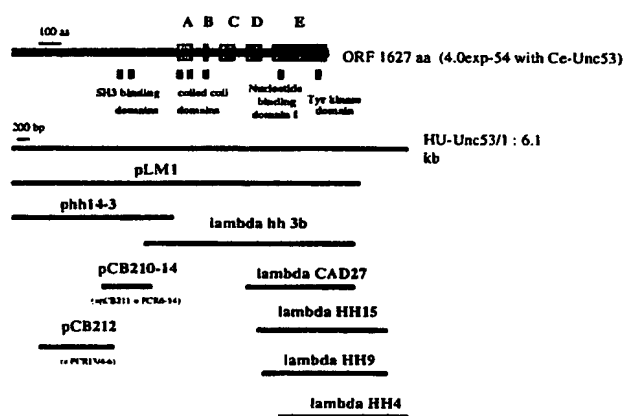


Figure 8a

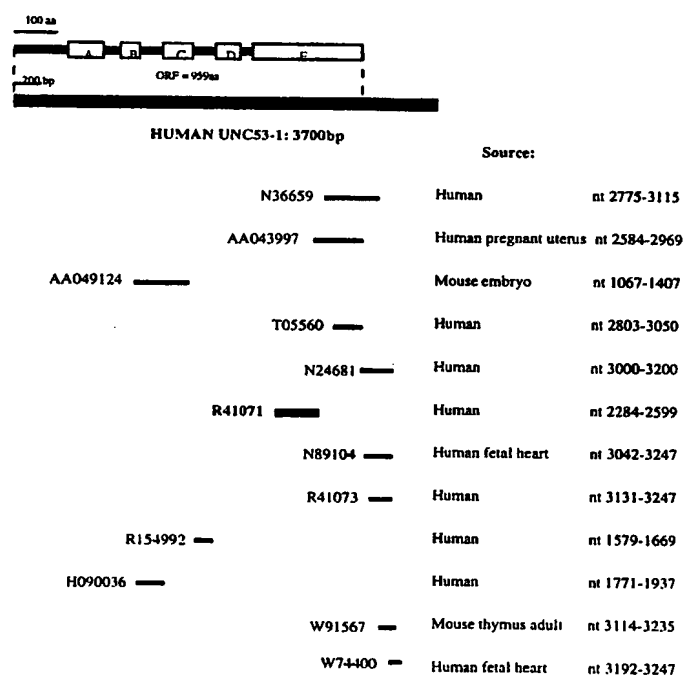


Figure 8b

FIG. 9a

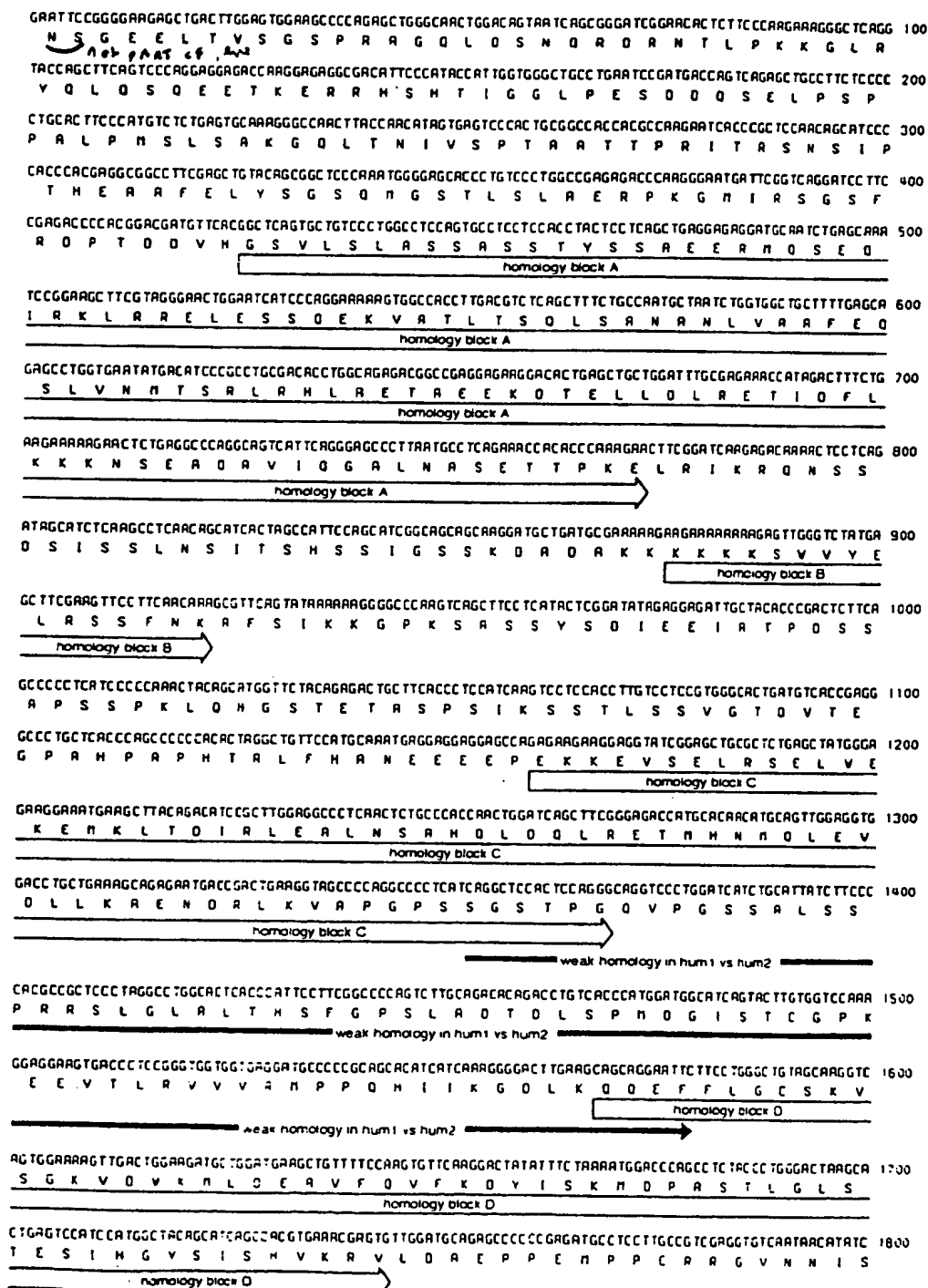


Fig. 9b

Tuesday, 18 November 1997 10:33

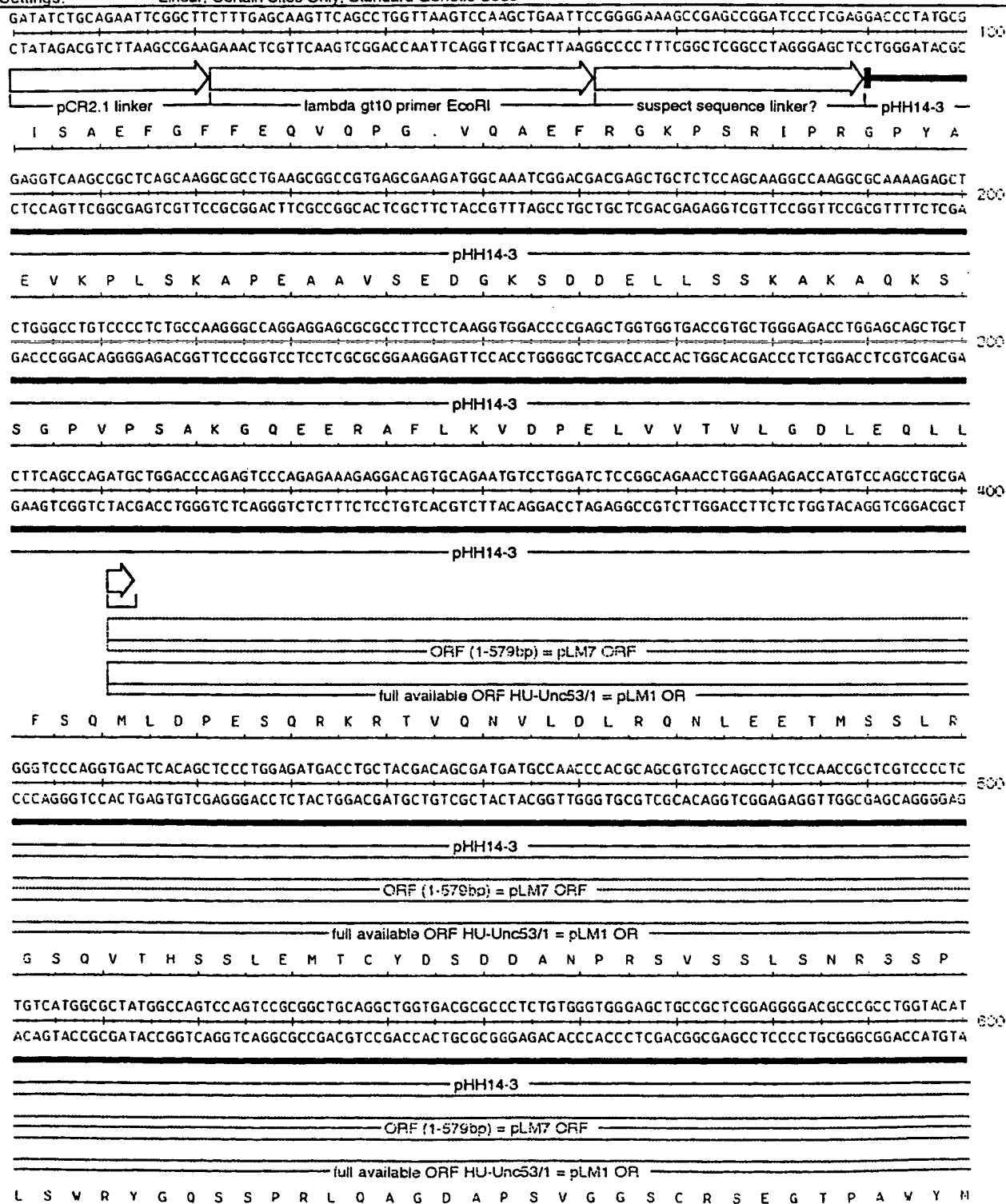
lu-Unc53/1 seq (1 > 6013) Site and Sequence

Enzymes: 60 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Seq 9 14 pag



Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 2

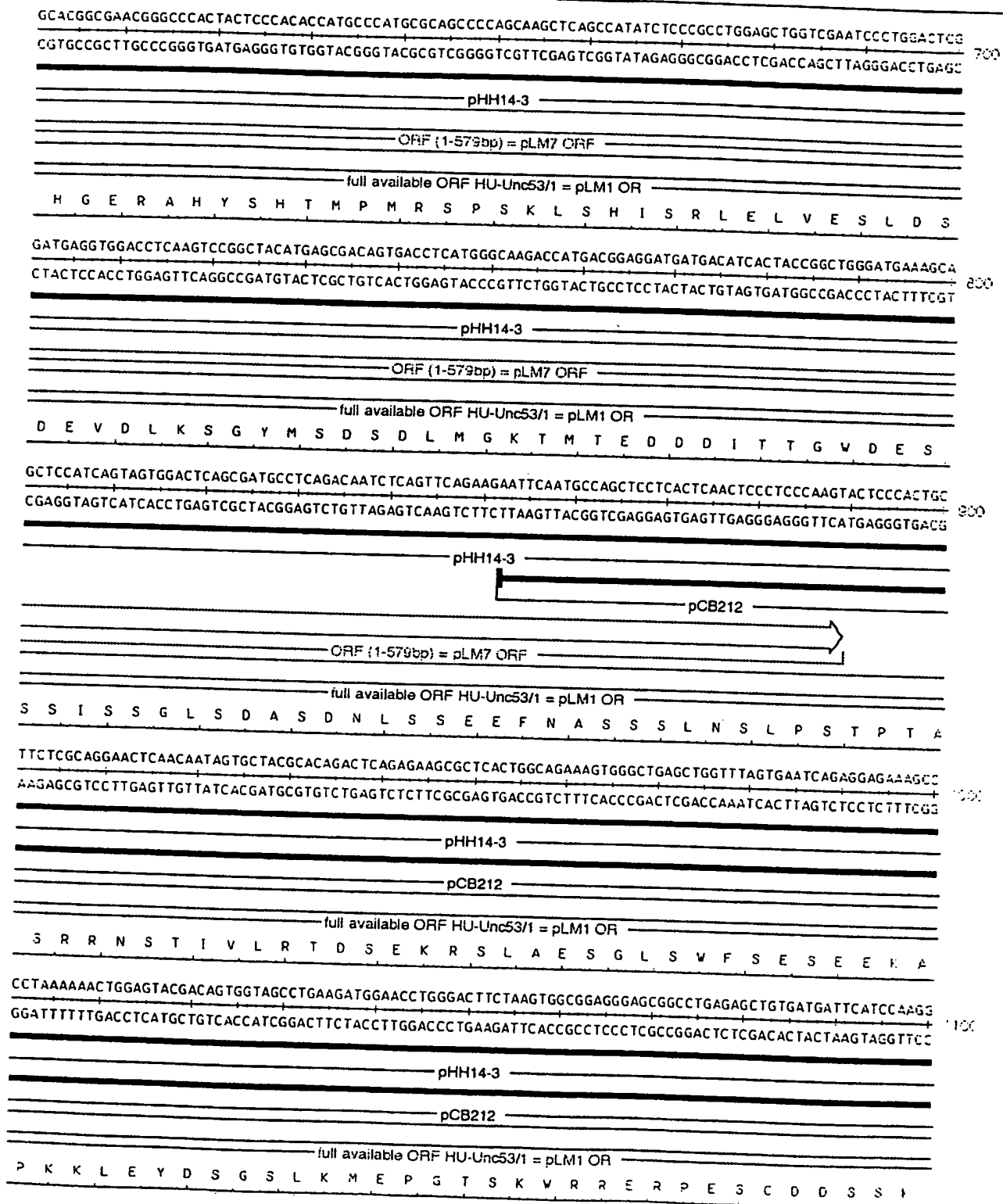


Fig 9b

Tuesday, 18 November 1997 10:33

Page 1

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCC TGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCCATCACTCACACAGC 120
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCG

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCCTCAAAGTCGCAGGCAAACTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCACTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT 130
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGTCTGTTCCTTCGAACTGACCTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

O S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA 140
CGACCAGCCC TGGCGGACTCACTACGATTCTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTCATGCAAACTGGTGGTTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC 150
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCTTCAGGAGTCCGTAGGACAGTTCGGTCATTACCCGCGTTCTG

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R F T

TAGCTTAGATGTTTCCAACAGTGCAAGCCAGGATTCTTGGCTCCTGGAGCCCCTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGT 160
ATCGAATCTACAAAGTTGTCACGTCCTGGTCTTAAGGACCGAGGACC TCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGGCCGGTTCGGTTCAGTTCA

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A K S S

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 3

TCTATGAGCGTGACCGGGGGGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCCTCAGCACCAAGCAGGGAGGCCCTTACGCCTTCCA
AGATACTCGCACTGGCCGCCCGCCACCTGGAGCGGGACACTCGTCGTCGTAAGTGGGGTCAGAGGAGTCGTGGTTCGTCCCCTCCGGAATGCGGAAGG 1700

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53rv4

S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC
CTGACTTCTCGGATGGTTCATCGGTCAACCGCTGGTGAGGTGCGGGACAGTTAGTCTGTCTAGCCCTTTTCTCTTCCGGTTTCGGTTCGGTCAACCG 1800

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N O T D R E K E K A K A V A

CTTGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTAC TCCAAGAACAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTG
GAACCTGAGTCTGTTGTAGAGGAAC TTCATAACGAGGGGTCTTTCATGAGGGTCTTGGTTCGTTCCGGTGGGGTGTGGTGGTTCGACCGTCTCGAC 1900

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S I G S P E S T P K N O A S H P T A T K L A E L

CCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTGTCAAACACCCCTACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCAT
GGTGGTGGGGAGAGTCCCGGTGTCGCTTCGAAACAGTTGGTGGGAGTGATCGGTTAGAAGTGTCCAGTTGAGGTTGTCAGACCTAGATGGTAGTA 2000

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L P S

Tuesday, 18 November 1997 10:33
fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 5

CCAGTGATACCAACCATGCTTCAAAGGTCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCCCTCCCTTCTGCTTACCCCCAGTCCGGCACC 210
GGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGTTCAGGCCGTGG

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P A P

CATCTCAATATTAACTCAGCCAGCTTCTCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACCTCTCAGGC 220
GTAGGAGTTATAATTGAGTCGGTCAAGAGGGTCCCGACCTCGATTACTACCAAAGTCACACGGTTTCTCTGGGCGTACATGGGGTTGAGAGTCCG

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L S G

CTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCCTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCTTCTGCTGCTC 230
GACGTGCTCTCGTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTACGGAAGGGTTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAG

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P A A

CCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTTCTCCCAAGAAAGG 240
GGTGTCTTCTTCTGCTTCTCGACTGAACCTACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTCTTTCC

pHH14-3

pHH3b

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53rv3

rev primer HU53rv2

peptide B72626H

P T E E E T E E L T W S G S P R A G Q L D S N O R D R N T L P K K G

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 6

GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTCGAAGTCAGGGTCCTCCTCTGGTTCTCTCCGTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGA 2600

pHH14-3

pHH3b

rev primer HU53/rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCAAGCAAGAATCACCCGCTCCAACA
AGAGGGGGACGTGAAGGTACAGAGACTCACGTTTCCCGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGGGTTCTTAGTGGGCGAGGTTGT 2600

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGCGGCCCTTCGAGCTGTACAGGGCTCCCAATGGGAGACACCTGTCCCTGGCCGAGAGACCAAGGGAATGATTGCGTCAGG
CGTAGGGGTGGGTGCTCCGCGGAAGCTCGACATGTGCGCGAGGGTTTACCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTAC TAAGCCAGTCC 2700

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G

ATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCTCCACCTACTCTCAGCTGAGGAGAGGATGCAATCT
TAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACCGAGGTACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGA 2800

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAATCCGGAAGCTTCGTAGGGAAC TGAATCATCCAGGAAAAAGTGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT
CTCGTTTAGGCCCTTCAAGCATCCCTTGACCTTAGTAGGGTCTTTTACCGGTGGAAGTCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA 2900

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

E Q I R K L R R E L E S S Q E K V A T L T S O L S A N A N L V A A

Fig 9b

Tuesday, 18 November 1997 10:33

fig HU-Unc53/1 seq (1 > 6013) Site and Sequence

Page 7

TTGAGCAGAGCC TGGTGAATATGACATCCCGCCTGCGACACC TGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGA
AACTCGTCTCGGACCACCTTATACGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCCTCTCCGTGACTCGACGACCTAAACGCTCTTTGGTATCT

300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTCAGGGAGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAC
GAAAGACTTCTTTCTTGAGACTCCGGGTCGTCAGTAAGTCCCTCGGAATTACGGAGTCTTTGGTGTGGGTTCTTGAAGCCTAGTTCCTGTTTTG

3100

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q H

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCTAGTATCGGTAAGGTCGTAGCCGTCGTCCTTACGACTACGCTTTTTCTCTTTTTTCTCAACCC

3200

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S W

TCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCAGTATAAAAAAGGGCCCAAGTCAGCTTCCTCATAC TCGGATATAGAGGAGATTGCTACACCCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGAAGTCATATTTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCCTCAACGATGTG36ET

3300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTACGCCCCCTCATCCCCAACTACAGCATGGTTCACAGAGACTGCTTCACCCCTCCATCAAGTCCTCCACCTTGTCTCCGTGGGCACTGATGT
GAGAAGTCGGGGGAGTAGGGGTTGATGTCTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGG TGGAACAGGAGGCACCCGTGACTACAG

3400

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q H G S T E T A S P S I K S S T L S S V S T D V

Fig 9b

Tuesday, 18 November 1997 10:33

Page 8

file: 'Hu-Unc53/1 seq (1 > 6013) Site and Sequence

ACCG .GCCCTGCTACCCAGCCCCACACTAGGCTGTTCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGC
TGGCTCCCGGACGAGTGGGTGCGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCTCCATAGCCTCGACGCGAGACTCG

3500

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGAGGAGCCCTCAACTCTGCCCACTGATCAGCTTCGGGAGACCATGCACAACATGCAGTT
ATACCCTCTTCTTTACTTCGAATGCTGTAGGCGAACCCTCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAA

3600

U2 ORF = pCB251 ORF

pHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCAC TCCAGGGCAGGTCCCTGGATCATCTGCATTA
CCTCCACCTGGACGACTTTTCGTCTCTTACTGGCTGACTTCCATCGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAA

3700

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCACGCCGCTCCCTAGGCTTGGCACTACCCATTCCTTCGGCCCCAGTCTTGACAGACAGACCTGTACCCATGGATGGCATCAGTACTTGTG
AGAAGGGGTGCGGCAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAACGTCTGTGTC TGGACAGTGGGTACCTACCGTAGTCATGAACAC

3800

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S T C

Tuesday, 18 November 1997 10:34

fig u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 9

GTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCGGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCTGGGGCTGTAS
CAGGTTTCTCTTCAC TGGGAGGCCACCACCACCTCTACGGGGCGTCGTGTAGTAGTTCCCTGAAC TTCGTCGTCCTTAAGAAGGACCCGACATC 330

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P O H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTC AAGTGTTC AAGGACTATATTTCTAAAATGGACCCAGCCTCTACCTGGGA
GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTC TGATATAAAGATTTTACCTGGGTCGGAGATGGGACCT 300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGACCTTGCTCACAACTACGTCTCGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT 310

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V H

ACATATCAGTCTCCCTCAAAGGTCTSAAGGAGAAATGCGTCGACAGCC TGGTGTTCGAGACGCTGATCCCCAAGCCGATGTCAGCACTACATAAGCCT
TGTATAGTCAGAGGGAGTTCCAGACTTCTCTTTACGCAGCTGTCGGACCAAGCTCTGCGACTAGGGGTTTCGGCTACTACGTGCTGATGTATTGGGA 320

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N I S V S L K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L

Tuesday, 18 November 1997 10:34
fig. 10-Unc53/1 seq (1>6013) Site and Sequence

Fig 9b

Page 1D

CCCTCTGAAGCACCGGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGCTGGAGCGGCTCTGGC
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTTCTGGATGGACTGGTTAGCGAACC GGCTCATGGACCACCTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACAGATAGACC
GCACTCCAGTGTCTCCCGTAGCAGTCTGGAAGTTGTACGTGGTCGTCAGAAGCTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAAACAGGAATTGGGGATGTGCCCCGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAA
CCCTTTGTCCCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGAGCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C I

Tuesday, 18 November 1997 10:34

fig. 4u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 17

GTATCA:AAATGTCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGACCTTGAGCTTCAGGATGTTGACCTTCCTCC
CATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAACGTGAACTCGAAGTCTTACAAC TGGAAAGG 480

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72626H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC
TTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCTTCGACCATCTCAGTCTGTGCGCTGTAGTTACGGTTGTTCCTTCTCGACGAAG 470

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCTTGAGAAGCACAGCACCTCAGACTTCTCATCGGCCCTTGCTTCTTCTGT
CCCACGAGCTGACCCATGGGTTGACACCATAGTAGAGGTGTGGAAGGAAC TCTCGTGTCTGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAG 460

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L D W V P K L V Y H L H T F L E K H S T S D F L : G P C F F L S

Tuesday, 18 November 1997 10:34

fig HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 12

GTGTCCCATTTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACCTGTGGAACAACCTATCATTCCCTATCTACAGGAAGGAGCCAAAGGATGGGATAAAG
CACAGGGTAACCGTAACCTCTGAAGGCC TGGACCAAGTAACCTGGACACCTTGTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCTACCTATTTCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCAAC
CAGGTACCTGTCTTTCGACGAACCCCTCTGGGTCACCTTACCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V H G Q K A A W E D P V E W V R D T L P W P S A Q Q D Q S K L Y H

TGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCCTCTGA
ACGGGGGTGGGTGGCACCCGGGAGTGTGTAACGGAGTGGAGGGCTCCTATCCTGTGAGTTTCTGTGCGTGGGTTCAAGAGACCTGAGTCTAGGAGACT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

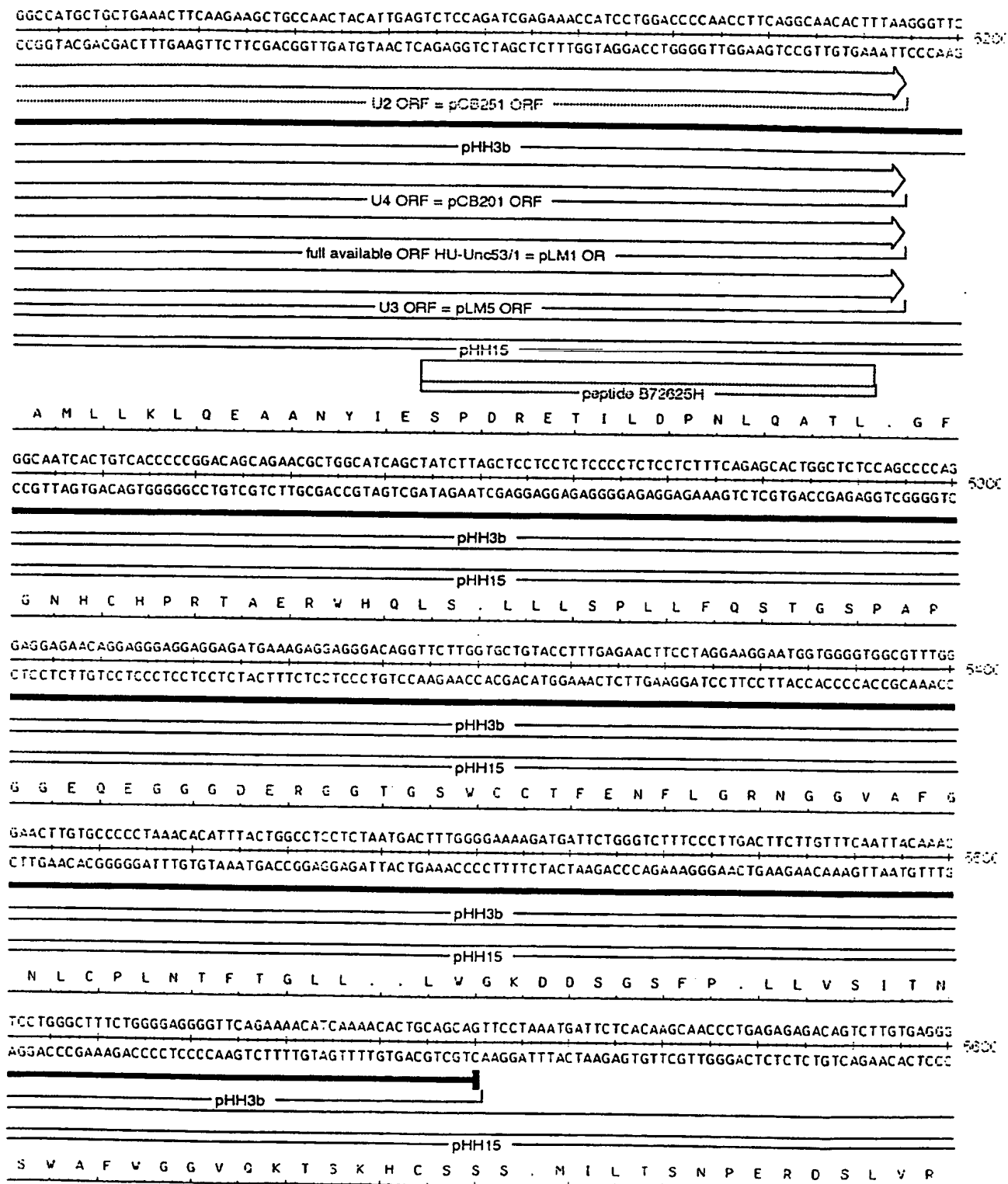
L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D P L I

Tuesday, 18 November 1997 10:34

fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 13



Tuesday, 18 November 1997 10:34
fig. Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 4

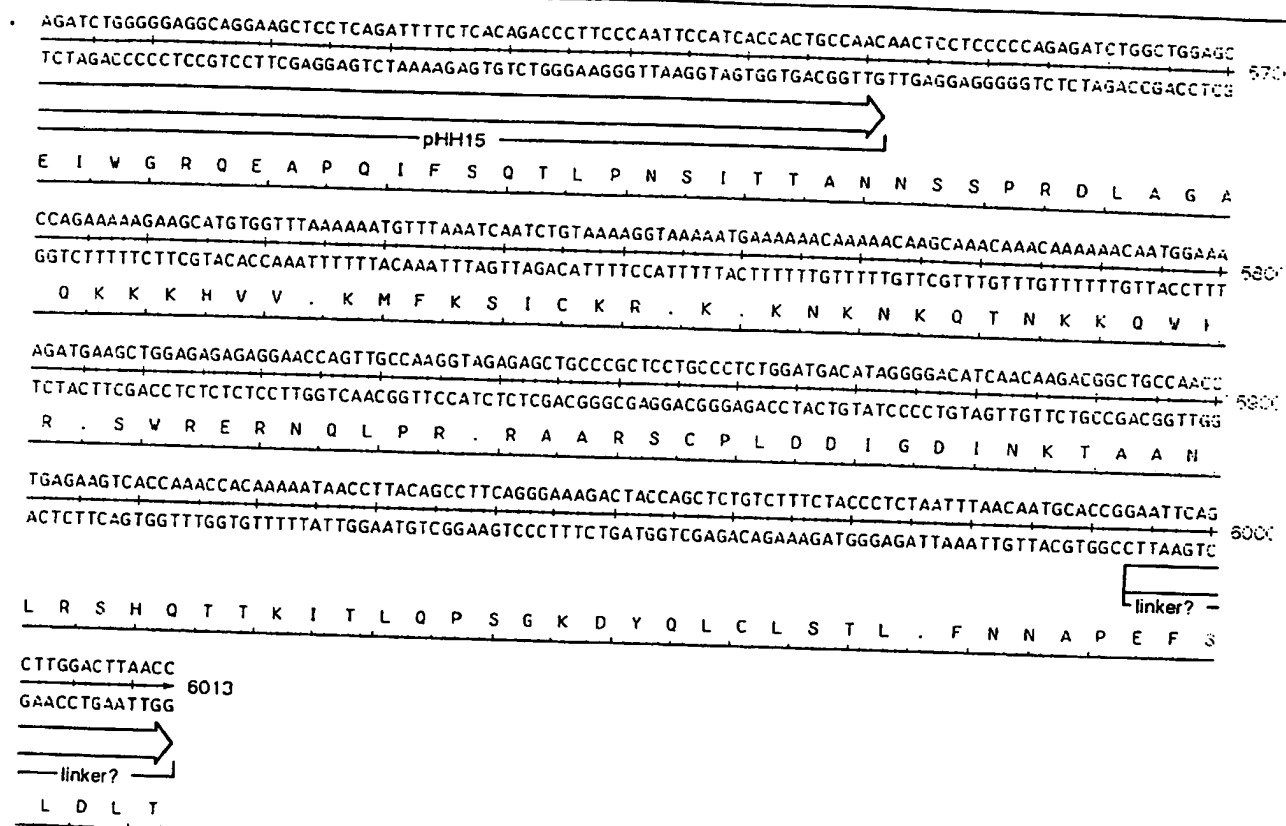


FIG. 10.

GGCACGAGGCATCCTCTGTGGGCACCGAGGTCACCGAGACCCCTGCTCATTCAGTCCCCACACTAGACT 70
linker ? open reading frame

R H E A S S V G T E V T E T P A H S V P H T R L
GTTCCAAGCCAATGAAGAGGAGGAGCCAGAGAAGAAGGAGGTATCAGAACTGCGCTCTGAACATATGGGAA 140
open reading frame

F O A N E E E E P E K K E V S E L R S E L V E
AAAGAGATGAAGCTCACGGATATCCGGTTGGAGGCCCTCAACTCTGCCACAGCTGGACCAGCTTCGGG 210
open reading frame

K E M K L T D I R L E A L N S A H O L D O L R
AGACCATGCACAATATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGGCTGAAGGTTGCCCC 280
open reading frame

E T M H N M O L E V D L L K A E N D R L K V A P
CGCCCCCTCTCAGGCTGCACTCCAGGGCAGGTCCCTGGGTCACTCGGCTCTGTCGTCCCCCTCGACGTTCC 350
open reading frame

G P S S G C T P G O V P G S S A L S S P R R S
CTGGGCCTTGCACTCAGCCATCCTTTTCAGTCCTAGTCTCACAGACACAGACCTCTCACCCATGGATGGCA 420
open reading frame

L G L A L S H P F S P S L T D T G L S P M D G
TCAGCACCTGTGGTTCAAAGGAAGAGGTGACCCCTCGGGTGGTGGTCCGGATGCCGCCCCAGCACATCAT 490
open reading frame

I S T C G S K E E V T L R V V V R M P P D H I
CAAAGGGGACTTAAAGCAGCAGGAGTTCTTCTGGGTTGCAGCAAGGTGAGTGGCAAAGTTGACTGGAAG 560
open reading frame

K G D L K Q Q E F F L G C S K V S G K V D K
ATGCTGGATGAAGCCGTTTTCCAAGTGTTCAAGSACTACATTTCTAAAAAGGAGCCAGCCTCAAGCCTGG 630
open reading frame

M L D E A V F Q V F D Y I S K M D E A S T L
GACTGAGCACTGAGTCCATACATGGCTATAGCCTCAGCCACGTGAACCGAGTGTGGAATGCTGAGTCCCC 700
open reading frame

G L S T E S I H G Y S L S H V K R V L D A E P

FIG. 10 CONTINUED.

AGAGATGCCCTCCCTGCCGCCGAGGTGTCAATAACATATCAGTCGCCTCAAAGGTCTGAAAGAGAAGTGT 770
open reading frame

E M P P C R R G V N N I S V A L K G L K E K C
GTCGACAGCCTGGTGTTCGAGACGCTTATCCCAAGCCCATGATGCAGCACTACATCAGCCTCCTGCTCA 840
open reading frame

V D S L V F E T L I P K P M M Q H Y : S L L L
AGCACCGGCGCCTGGTGCTCTCCGGCCCCAGTGGCACCAGGCAAGACCTACTTGACCAATCGGCTAGCCGA 910
open reading frame

K H R R L V L S G P S G T G K T Y L T N D L A E
GTACCTGGTGGAGCGCTCCGGCCGCGAGGTACGGATGGCATCGTCAGCACTTTCAACATGCACCAGCAG 980
open reading frame

Y L V E R S G R E V T D G I V S T F N M H Q Q
TCTTGCAAGGATCTGCAACTGTACCTCTCCAACCTAGCCAACCAGATAGACCGGGAACAGGGATAGGGG 1050
open reading frame

S C K D L Q L Y L S N L A N Q I D R E T S I S
ATGTGCCCTTGGTGATCCTCCTGGATGATCTGAGTGAAGCAGGCTCCATCAGTGAGCTGGTCAATGGGGC 1120
open reading frame

D V P L V I L L D D L S E A G S I S E L V N G A
CCTCACCTGCAAGTATCACAATGTCCCTACATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCC 1190
open reading frame

L T C K Y H K C P Y I I G T T N O P / K M T P
AACCATGGCTTGACCTTGAGCTTCAGGATGCTGACCTTCTCGAACAATGTGGAACCAGCCAATGGCTTTC 1260
open reading frame

N H G L H L S F R M L T F S N N V E P A N G F
TGGTCCGTTACCTGCGGAGGAAGTTGGTAGAGTCAGACAGTGACGTCAATGCTAACAGGAGAGCTGCT 1330
open reading frame

L V R Y L R R K L V E S D S D V N A N E E L L
TCGGGTGCTGGACTGGGTGCCCAAGCTGTGGTATCACCTCCACACCTTCCTGGAGAAGCAAGCACCTCG 1400
open reading frame

R V L D V V P K L V H L H T F L E H S T S

GACTTCCTCATTGGCCCTTGCTTCTTCCTGTCTGTCCATTGGCATCGAGGACTTCCGGACCTGGTTCA 1470
open reading frame

D F L I G P C F F L S C P I G I E D F R T V F
TTGACCTGTGGAACAATTCCATCATCCCCTATCTACAGGAAGGAGCCAAGGATGGGATCAAGGTTTCATGG 1540
open reading frame

I D L W N N S I I P Y L Q E G A K D G I K V H G
ACAGAAAGCTGCTTGGAAGACCCGGTGGAATGGGTCCGAGACACTCTTCCTGGCCGTCGGCCCAACAA 1610
open reading frame

Q K A A V E D P V E W V R D T L P W P S A Q Q
GACCAATCAAAGCTCTACCACCTGCCCCCGCCTTCTGTGGGCCCCACAGCACTGCCTCACCCCGGAGG 1680
open reading frame

D Q S K L Y H L P P P S V G P H S T A S P P E
ACAGGACAGTCAAAGACAGCACTCCAAACTCCCTCGAC TCAGATCCCTTGATGGCCATGCTACTGAAACT 1750
open reading frame

D R T V K D S T P N S L D S D P L M A M L L K L
CCAAGAAGCTGCCAACTACATTGAGTCACCAGATCGAGAGACTATCCTGGACCCCAACCTCCAGGCGACA 1820
open reading frame

Q E A A N Y I E S P D R E T I L D P N L Q A T
CTCTGAGGGCCCCGGCAGTCACTGTACCCTGGAGGGCAGAAGGCTGGCTTCAGCATCATTAGCTCTCCTC 1890
3' untranslated

L . G P G S H C H P G G O K A G F S I I S S P
TGCCCTCTTCCTTCATAGCTCTGGCTCACCAGCCTCGCCAAGAGAACAGGAGGGAAGAAGAGGGCAGGAG 1960
3' untranslated

L P S S F I A L A H Q P R Q E N R E E E E G R R
GAGGGATGGGTCTCGGTGCTGAACCTTTGAGAACTTCCTACTAGGAATTGGAGGGGGTGGAGTTTGAGA 2030
3' untranslated

R D G F S V L N L . E L P T R N W R G V S L R
ACTCCGTGCCCTTAACACATTTCCTGGCCCTCTCTTACGACTTAGGAGAAAGATGATCTGGTCTTT 2100
3' untranslated

T P C P L T T F A G L L L R L R R K D D S G L
TCTTCAAGTTTGTTCACCTACAAACTCTTGGGCTTCTTGGGGAGGGATTCGGAAGATATAAACAGACA 2170
3' untranslated

F F K F C F T Y K L L G F L G R D S E D I N P D

FIG. 10 CONTINUED.

AACAAAAACAAACAACTACAGCAGTTCCAAGCTCGTTCTCACAACACCTCTGAGACAGTCACAT 2240
3' untranslated
T K T N K P T T A V P S S F S O T P L R Q S H
TGGGGCAAATCTAAGGGAGGCAGGAAGCTCTACAGACTTTCTTGCAAACCTTCCCAGTTCTGTGACAC 2310
3' untranslated
V G K S K G G R K L Y R L S C K P F P V L S T
TGCCAACAACCTCCCCGCCAGAGACCTGGCCAGAGCCAAGAAAAGAGAAGCATGTGGTTTAAACAGAAAAA 2380
3' untranslated
L P T T S P P E T V P E P R K E K H V V . Q K N
CAAAACAAAACAAAACAAAATATATGTGTAAATCAACCTGTAGAAGGTAAAAACGGCAATGGAAAAA 2450
3' untranslated
K T K Q N K K Y M C K S T C R R . K R Q V K R
TGAAGCTGGAAGGAGGGGCCAGTTGCCAAGATGGAACGAGAGCTGCCAGATCTTGCCCTCTGGATGACA 2520
3' untranslated
. S V K E G P S C Q D G T R A A R S C L L D D
AGAGGGGACATTGCAAGATGGCTGCCAGTCTAAAACGTCACCAGACCACAAGAGTAACATCAGCCTTC 2590
3' untranslated
K R G H C K M A A S L K R H Q T T R V T S O P S
GAAGAAAGGCCACAGCTGTCTTTCTGCCCTCTAACTGAACATGCATGAAAAGTCAATAAACCTACTTT 2660
3' untranslated
K K G H K L S F C P L T E H A . K V N K P Y F
TTAATTTTTAAAAAATTTCCGCGGCCGC 2709
polyA tail + linker
L I F K K K K K K K K F P R P

FIG. 11a.

AAGCTTGGCAGGAGGCTCGTGCCAAGCTGAGACCGTCATGCAGCTCCGAAATGAGTTAAGAGACAAGGA 70
LINKER ? open reading frame
A W H E A S C Q A E T V M Q L R N E L R D K E
GATGAAGCTGACAGATATCCGCTTAGAAGCTCTCAGTTCTGCCACCAGCTGGACCAGCTCCGGGAGGCC 140
open reading frame
M K L T D I R L E A L S S A H Q L D Q L R E A
ATGAACAGGATGCAGAGTGAAATAGAGAAGCTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAG 210
open reading frame
M N R M Q S E I E K L K A E N D R L K S E S Q
GCAGTGGCTGCAGCCGGGCTCCTTCCCAAGTGTCCATCTCTGCCTCCCCGAGGCAGTCCATGGGCTCTC 280
open reading frame
G S G C S R A P S Q V S I S A S P R Q S M G L S
CCAGCACAGCTTGAACCTCACTGAGTCAACCAGCCTGGACATGTTGCTGGATGACACTGGTGAATGCTCC 350
open reading frame
Q H S L N L T E S T S L D M L L D D T G E C S
GCTCGGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTCAGCTTTCAGGAGGAAATGAAGTGAAGGAGG 420
open reading frame
A R K E G G R H V K I V V S F Q E E M K V K E
ATTCCAGACCACACCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGAAGACGAAGTGGATGTGCTCGA 490
open reading frame
D S R P H L F L I G C I G V S G K T K V D V L D
TGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCATGTCGACCCAGTGAGTCAGCTAGGGCTGAAT 560
open reading frame
G V V R R L F K E Y I I H V D P V S Q L G L N
TCAGACAGCGTTCTTGGCTACAGCATTGGAGAAATCAAGCGCAGCAACACTTCCGAAACACCGGAGCTGC 630
open reading frame
S D S V L G Y S I G E I K R S N T S E T P E L
TTCCTTGTGGCTATCTGGTTGGAGAGAACACGACCATCTCAGTGACTGTGAAAGGGCTCGCAGAAAACAG 700
open reading frame
L P C G Y L V G E N T T I S V T V K G L A E N S
CCTGGACTCACTGGTGTGAGTCTTGATTCCCAAGCCATCCTGCAGCGCTACGTCTCCCTCCTGATA 770
open reading frame
L D S L V F E S L I P K P I L Q R Y V S L L :
GAGCACCGTCGGATCATCTCTTGCCCCAGCGGCAC TGGGAAAACCTACCTGGCCAACCGGCTGTCTG 840
open reading frame
E H R R : I L S G P S G T G K T Y L A N R L S

FIG. 11a CONTINUED.

AGTATATAGTGCCTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGCCACCTTTAACGTGGACCATAA 910
open reading frame
E Y I V L R E G R E L T D G V I A T F N V D H K
GTCCAGCAAGGAATTGCGCCAGTACCTGTCCAACCTTGCTGACCAGTGCAACAGTGAGAACAATGCTGTG 980
open reading frame
S S K E L R O Y L S N L A D Q C N S E N N A V
GACATGCCCTCGTCATCATCCTGGACAACCTACACCACGTGAGCTCTCTGGGCGAGATCTTCAATGGGC :050
open reading frame
D M P L V I I L D N L H H V S S L G E I F N G
TGCTCAACTGCAAGTACCACAAATGCCCTTACATAATTGGCACAATGAACCAGGCTACCTATCTCCCCCT :120
open reading frame
L L N C K Y H K C P Y I I G T M N Q A T Y L P F
TTATACTAATAATCTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTCTGCGGCCGC :190
open reading frame
LINKER-vector
Y T N N L I K K K K K K K K K K K K K K K F C G R

FIG. 11b.

		10	20	30	40	50
1 1.. 50 UNC53_huma	AAGCTTGGCA CGAGGCCTCG TGCCAAGCTG AGACCGTCAT GCAGCTCCGA					
5 1.. 48	A W M E A S C Q A E T V M Q L R					
	60 LINKER > 70 ORF (START) 80 90 100					
1 51..100 UNC53_huma	AATGAGTTAA GAGACAAGGA GATGAAGCTG ACAGATATCC GCTTAGAAGC					
5 49.. 98	N E L R D K E M K L T D I R L E A					
	110 120 130 140 150					
1 101..150 UNC53_huma	TCTCAGTTCT GCCCACCAGC TGGACCAGCT CCGGGAGGCC ATGAACAGGA					
5 99..148	L S S A H Q L D Q L R E A M N R					
	160 170 180 190 200					
1 151..200 UNC53_huma	TGCAGAGTGA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA					
5 149..198	M Q S E I E K L K A E N D R L K S					
	210 220 230 240 250					
1 201..250 UNC53_huma	GAGTCTCAAG GCAGTGGCTG CAGCCGGGCT CTTCCCAAG TGTCATCTC					
5 199..248	E S Q G S G C S R A P S Q V S I S					
	260 270 280 290 300					
1 251..300 UNC53_huma	TGCCTCCCGG AGGCAGTCCA TGGCCCTCTC CCAGCACAGC TTGAACCTCA					
5 249..298	A S P R Q S M G L S Q H S L N L					
	310 320 330 340 350					
1 301..350 UNC53_huma	CTGAGTCAAC CAGCCTGGAC ATGTTGCTGG ATGACACTGG TGAATGCTCG					
5 299..348	T E S T S L D M L L D D T G E C S					
	360 370 380 390 400					
1 351..400 UNC53_huma	GCTCCGAAGG AAGGAGGCAG GCATGTTAAG ATAGTTGTCA GCTTTCAGGA					
5 349..398	A R K E G G R H V K I V V S F Q E					
	410 420 430 440 450					
1 401..450 UNC53_huma	GGAAATGAAG TGAAGGAGG ATTCCAGACC ACACCTCTTT CTATTGGCT					
5 399..448	E M K W K E D S R P H L F L I G					
	460 470 480 490 500					
1 451..500 UNC53_huma	GCATTGAGT TAGTGGCAAG ACGAGTGGG ATGTGCTCGA TGGGGTGGTT					
5 449..498	C I G V S G K T K W D V L D G V V					
	510 520 530 540 550					
1 501..550 UNC53_huma	AGACGGCTGT TCAAAGAATA CATCATTCAT GTGACCCAG TGAGTCAGCT					
5 499..548	R R L F K E Y I I H V D P V S Q L					
	560 570 580 590 600					
1 551..600 UNC53_huma	AGGGCTGAT TCAACAGCG TTCTTGGCTA CAGCATTTGA GAAATCAAGC					
5 549..598	G L N S D S V L G Y S I G E I K					
	610 620 630 640 650					
1 601..650 UNC53_huma	GCAGCAACAC TTCCGAACA CCGGAGCTCC TTCTTSTGG CTATCTGGTT					
5 599..648	R S N T S E T P E L L P C G Y L V					
	660 670 680 690 700					
1 651..700 UNC53_huma	GGAGAGAACA CGACCATCTC AGTGAATCTG AAAGGGCTTC CAGAAAACAG					
5 649..698	G E N T T I S V T V K G L A E N S					

FIG. 11b CONTINUED.

701..750 UNC53_huma 5 699..748	710 720 730 740 750 CCTGGACTCA CTGGTGTTC AGTCCTTGAT TCCCAAGCCC ATCCTGCAGC L D S L V F E S L I P K P I L Q
751..800 UNC53_huma 5 749..798	760 770 780 790 800 GCTACGCTTC CCTCTGATA GAGCACCCTC GCATCATCTC CTCTGGCCCC R Y V S L L I E H R R I I L S G P
801..850 UNC53_huma 5 799..848	810 820 830 840 850 AGCGGCACTG GGAAAACCTA CCTGGCCAC GGGCTCTCTC AGTATATAGT S G T G K T Y L A N R L S E Y I V
851..900 UNC53_huma 5 849..898	860 870 880 890 900 GCTCGAGAG GGACGGGAGT TGACAGACGG GGTATCGCC ACCTTTAAGC L R E G R E L T D G V I A T F N
901..950 UNC53_huma 5 899..948	910 920 930 940 950 TGGACCATAA GTCCAGCAAG GAATTGGCC AGTACCTCTC CAACCTTGCT V D H K S S K E L R Q Y L S N L A
951..1000 UNC53_huma 5 949..998	960 970 980 990 1000 GACCACTGCA ACAGTGAGAA CAATGCTGTG GACATGCCCC TGTTCATCAT D Q C N S E N N A V D M P L V I I
1001..1050 UNC53_huma 5 999..1048	1010 1020 1030 1040 1050 CCTGGACAAC GTACACCAG TGAGCTCTCT GGGCGAGATC TTCAATGGCC L D N L H H V S S L G E I F N G
1051..1100 UNC53_huma 5 1049..1098	1060 PRIMER 1070 1080 1090 1100 TGTCTCACTG CAAGTACCAC AAATGCCCTT ACATAATGG CACAATGAAC L L N C K Y H K C P Y I I G T M N
1101..1150 UNC53_huma 5 1099..1148	1110 1120 1130 1140 1150 CAGGCTACCT CTTCGACTCC CACCTCCAG CTTCACCATA ACTTCAGATG C A T S S T P N L Q L H H N F R W NOVEL SEQUENCE NOT IN Fig 11a
1151..1200 UNC53_huma 5 1149..1198	1160 1170 1180 1190 1200 GGTGCTTTGT GCCAACCACA CCGAGCCTGT GAAGGGTTTC CTGGCCGAT V L C A N H T E P V K G F L G R
1201..1250 UNC53_huma 5 1199..1248	1210 1220 1230 1240 1250 TCCTGAGGAG GAAGCTCATG GAAACAGAGA TCAGTGGGCS GGTGGCCAAT F L R R K L M E T E I S G R V R N
1251..1300 UNC53_huma 5 1249..1298	1260 1270 1280 1290 1300 ATGGAGCTGG TAAAAATCAT TGACTGGATT CCAAGGTCT GGCATCACCT M E L V K I I D W I P K V W H H L
1301..1350 UNC53_huma 5 1299..1348	1310 1320 1330 1340 1350 CAACCCCTTC CTGGAGGCTC ACAGTTCCTC GACGTCACC ATCGGCCCCC N R F L E A H S S S D V T I G P
1351..1400 UNC53_huma 5 1349..1398	1360 1370 1380 1390 1400 GCTCTCTCCT GTCATGCCCC ATCGATCTCG ACGGCTCCAG AGTGTGCTTC R L F L S C P I D V D G S R V W F

FIG. 11b CONTINUED.

		1410	1420	1430	1440	1450
1 1401...1450 UNCS3_hu		ACCGACTTGT	GGAACATATC	CATTATCCCC	TATCTCCTGG	AAGCCGTCAG
5 1399...1448		T D L W N Y S	I I F Y L L	E A V R		
		1460	1470	1480	1490	1500
1 1451...1500 UNCS3_hu		AGAAGGACTC	CAGCTCTATG	GAAGGCGCGC	CCCCTGGGAG	GATCCTGCCA
5 1449...1498		E G L Q L Y	G R R A	P W E D P A		
		1510	1520	1530	1540	1550
1 1501...1550 UNCS3_hu		ACTGGGTGAT	GGACACATAT	CCATGGGCGAG	GCAGGCCACA	ACAGCAGCAG
5 1499...1548		K W V M D T Y	P W A A S P Q	Q H E		
		1560	1570	1580	1590	1600
1 1551...1600 UNCS3_hu		TGGCCTCCGC	TGCTGCAGTT	ACGGCCTGAG	GATGTGGGCT	TGACGGGTA
5 1549...1598		W P P L L Q L	R P E D V G	F D G Y		
		1610	1620	1630	1640	1650
1 1601...1650 UNCS3_hu		CTCCATGCCT	CGGAGGGGAT	CGACAAGCAA	GCAGATGCCC	CCAGTGATG
5 1599...1648		S M P R E G	S T S K Q M P	P S D		
		1660	1670	1680	1690	1700
1 1651...1700 UNCS3_hu		CTGAAGGTGA	CCGCTGATG	AACATGCTGA	TGAGGCTGCA	GGAGGCAGCC
5 1649...1698		A E G D P L M	N M L H R L Q	E A A		
		1710	1720	1730	1740	1750
1 1701...1750 UNCS3_hu		AATCACTCCA	GCCCCAGAG	CTATGACAGC	GACTCCAACA	GCACAGCCA
5 1699...1748		N Y S S P Q S	Y D S C S N	S N S H		
		1760	1770	1780	1790	1800
1 1751...1800 UNCS3_hu		TCACGATGAC	ATCTTGGACT	CCTCTTTGSA	GTCACCTGTG	TGACAGGGGC
5 1749...1790		H D D I L D	S S L E S T L	ORF		
		1810	1820	1830	1840	1850
1 1801...1850 UNCS3_hu		CCGGAGCTCA	GCGCCCTCCT	CTTCTCTCA	CCGCACTCA	CCTGCATCCC
5 ---,---		<==				
		1860	1870	1880	1890	1900
1 1851...1900 UNCS3_hu		CACATCACCC	TGAAGATGAC	TTCCTGAGCC	AGCCCCAGCC	ACAGCCTTAG
5 ---,---		<==				
		1910	1920	1930	1940	1950
1 1901...1941 UNCS3_hu		AGCTGGGGA	ACACCGAGAC	CCCCGTCTT	CAGCCTGAC	T
5 ---,---		<==				

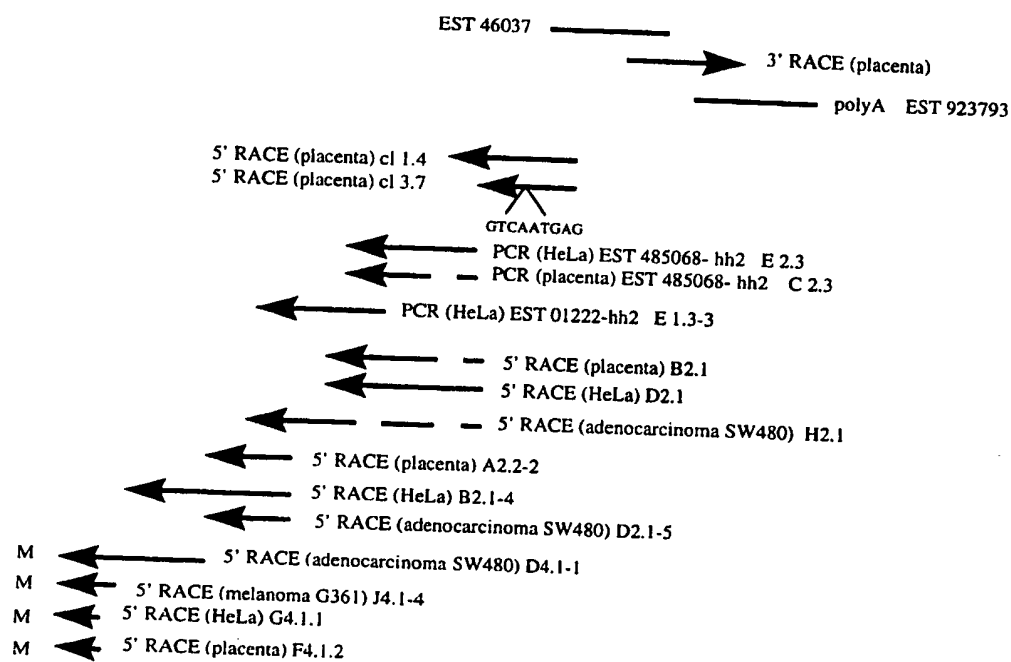


Figure 11c

[illegible]

+1 T E P V K G F L G R F L R R K L M E T E I S G R V R N N
 CGACGCTGOT GAAGGCTGTC CTGCTGAGG TACTGAGG GAGACGAG TACTGCTGCG GGTGCTGCG
 CGCTCGGAGA CTCTCGGAGG AGAGCTCTC GTCCTGAGT CTCTGCTCT ACTGACGCG CGACGGTTA
 64601
 +1 M E L V K I I D W I P K V W H H L N R F L E A H S S
 TACTGCTGAC TAAMATCTAT TACTGATAT CGATCTGAT GCGATCTCT CAGCTGCTC CTGAGCTGTC AGCTGCTCT
 64681
 +1 D V T I G P R L F L S C P I D V D G S R V W F T D L
 GAGCTGCTGCG CGAGCTGCTG GCGATCTGOT GTCATGCTG ATGCTGCG AGGCTGCG ACTGCTGCTC AGCGATCTCT
 64683
 +1 M N Y S I I P Y L L E A V R E G L Q L Y G R R A P W E
 GAGACTCTC CATATCTGCG TACTGCTGCG ATGCGCTGAG AGAGAGCTC GAGCTGATCT GAGGCTGCG CGCTGCTGCG
 CTCTGAGG GTATATGCG ATAGAGCGC TCGCGCTGTC TCTCTGAGT CTGTGATGAC CTCTGCTGCG GAGGACCTG
 64684
 +1 D P A C K W V H D T Y P W A S P Q Q H E W P P L L Q
 CTCTGCTGCA ATGCGCTGCT GAGACTATY CTATGCTGCG GAGCTGCTGCA AGACGCTGCA TCGCTGCTGCT TCGCTGATCT
 64721
 +1 A G C E T G V D G F D G Y S W P R E G S T S K Q M P P S D
 TCGGCTGCTC CTACGCTGCA ACTCTGATAT GAGCTGCTGCA GAGCTGCTGCT TCTGCTGCTC AGCTGCTGCG GCGCTGCTAC
 64801
 +1 A E G D P L M N M L H R L Q E A A N Y S S P O S Y D S
 CTGAGCTGCA CGCTGCTGCT ACATCTGCA TACGCTGCA GAGGCTGCT ACTGCTGCTC TCTGCTGCTC TCTGCTGCTC CTATGCTGCTC
 64881
 +1 D S N S N S H E D I L D S S L E S T L Q G P G A
 GAGCTGCTGCT CTCTGCTGCT TACGCTGCA ATCTGCTGCT CTCTGCTGCT GTCATCTGCT TACGCTGCTGCT CGGAGCTGCT
 64961
 +1 R P P L L L T A F H L H P P H P E D D F L S Q P P
 CGCTGCTGCT CTCTGCTGCT CGCTGCTGCT CTGCTGCTGCT CGACATCTGCT CTGATGCTGCT GAGCTGCTGCT CTGCTGCTGCT
 7041
 +1 A T A L E I R E H R D P P S F S L D L G A G I P G Q L
 CGAGCTGCT AGCTGCTGCG GAGACGCTGCT ATCTGCTGCT CTCTGCTGCT GAGCTGCTGCT GAGCTGCTGCT CTGCTGCTGCT
 7121
 +1 P A D R F L P Q R E L H Y L L L Y F N Y C F A L L
 CTGCTGCTGCT GCTCTGCTGCT ACAGCTGCA CTGCTGCTGCT TCTGCTGCT CTCTGCTGCT CTCTGCTGCT TCTCTGCTGCT
 7201
 +1 P P D T E D T S R E R I A V E M K K K K K K
 ACTGCTGCTA GACATGAG TACTGCTGCG GAGGAGCTC ATCGCTGCT ATATGAGCTA ATATATATAT ATATATATAT ATATATATAT
 7281
 +1 K K K K
 ATATATATAT AT
 ATATATATAT

[illegible]

11/24/97 02:33PM

Fig 11e

EST 923793

Page 1

+1 G T R V T I G P R L F L S C P I D V D G S R V W F T D
1 GGCACGAGGG TTACCATCGG CCCCCGGCTC TTCCTGTCTAT GCCCATCGA TGTGGACGGC TCGAGAGTGT GGTTCACCGA
CCGTGCTCCC AATGGTAGCC GGGGGCCGAG AAGGACAGTA CGGGGTAGCT ACACCTGCCG AGCTCTCACA CCAAGTGGCT
.....
+1 L W N Y S I I P Y L L E A V R E G L Q L Y G R R A P
81 CTTGTGGAAC TATTCCATTA TCCCCTATCT CCTGGAAGCC GTCAGAGAAG GACTCCAGCT CTATGGAAGG CGCGCCCCCT
GAACACCTTG ATAAGTAAT AGGGGATAGA GGACCTTCGG CAGTCTCTTC CTGAGGTGCA GATACCTTCC GCGCGGGGGA
.....
+1 W E D P A K W V M D T Y P W A A S P Q Q H E W P P L L
NcoI

161 GGGAGGATCC TGCCAAGTGG GTGATGGACA CATATCCATG GGCAGCCAGC CCACAACAGC ACGAGTGGCC TCCCCTGCTG
CCCTCCTAGG ACGGTTCAAC CACTACCTGT GTATAGGTAC CCGTCGGTCG GGTGTTGTCG TGCTCACCGG AGGGGACGAC
.....
+1 Q L R P E D V G F D G Y S M P R E G S T S K Q M P P S
241 CAGTTACGGC CTGAGGATGT CGGCTTCGAC GGCTACTCCA TGCCTCGGGA GGGATCGACA AGCAAGCAGA TGCCCCCAG
GTCAATGCCG GACTCCTACA GCCGAAGCTG CCGATGAGGT ACGGAGCCCT CCCTAGCTGT TCGTTCGTCT ACGGGGGGTC
.....
+1 D A E G D P L M N M L M R L Q E A A N Y S S P Q S Y
321 TGATGCTGAA GGTGACCCGC TGATGAACAT GCTGATGAGG CTGCAGGAGG CAGCCAACTA CTCCAGCCCC CAGAGCTATG
ACTACGACTT CCACTGGGCG ACTACTTGTA CGACTACTCC GACGTCTCTC GTCGGTTGAT GAGGTGCGGG GTCTCGATAC
.....
+1 D S D S N S N S H H E D I L D S S L E S T L * Q G P G
401 ACAGCGACTC CAACAGCAAC AGCCATCAGC AAGACATCTT GGACTCCTCT TTGGAGTGCA CTCTGTGACA GGGGCCCCGA
TGTGCTGAG GTTGTGCTG TCGGTAGTGC TTCTGTAGAA CCTGAGGAGA AACCTCAGGT GAGACACTGT CCCCCGGCT
.....
+1 A Q R P P L L L T A F H L H P P H H P E D D F L S Q P
481 GCCCAGCGCC CTCTCTTCT CCTCACCGCA TTCCACCTGC ATCCCCCACA TCACCCGTGA GATGACTTCC TGAGCCAGCC
CGGTGCGCG GAGGAGAAGA GGAGTGGCGT AAGGTGGACG TAGGGGGTGT AGTGGGACTT CTACTGAAG ACTCGGTGCG
.....
+1 P A T A L E L R E H R D P P S F S L D L G A G I P G
561 CCCAGCCACA GCCTTAGAGC TCGGGGAACA CCGAGACCCC CCGTCTCTCA GCCTCGACCT GGTGCGAGGC ATCCCGGGCC
GGGTGCGTGT CGGAATCTCG ACGCCCTTGT GGCTCTGGGG GGCAGGAAGT CGGAGCTGGA CCCACGTCCT TAGGGCCCCG
.....
+1 Q L P A D R F L P Q R E L H Y L L L Y F N Y C F A L L
641 AGCTGCCTGC GGACCGCTTCT CTTCCACAGC GAGAACTGCA CTACCTTCTG TTGTACTTTA ATTATTGTTT TGCCTTGTTG
TCGACGGACG CTTGGCGAAG GAAGGTGTCT CTCTTGACGT GATGGAAGAC AACATGAAAT TAATAACAAA ACGGAACAAC
.....
+1 L * P P * D T E D T S R E R I I A V E M K K K K K K K
721 CTGTGACCTC CTAAGACAC TGAAGATACT TCTCGGGAAA GGATCATCGC CGTTGAAATG AAAAAAAAAA AAAAAAAAAA
GACTGAGG GGATTCTGTG ACTTCTATGA AGAGCCCTTT CTTAGTAGCG GCAACTTTAC TTTTTTTTTT TTTTTTTTTT
.....
+1 K K K K K N E G G R K L
801 AAAAAAAAAA AAAAAAACG AAGGCGCGC CAAGCTT
TTTTTTTTTT TTTTTTTTGC TTCCGCGGC GTTCGAA

11/24/97 02:40PM

hh2UNC53_5'Race_c11.4_CON_REV.seq

Page 1

+1 I P L T I G L E R P P R Q V P H L D R N T L P K K G L
1 ATCCCACTCA CTATAGGGCT CGAGCGGCGG CCCAGGCAGG TCCGCACCT TGATAGGAAC ACTTTGCCTA AGAAAGGACT
TAGGGTGAGT GATATCCCGA GCTCGCCGCG GGGTCCGTCC AGGGCGTGGA ACTATCCTTG TGAAACGGAT TCTTTCCTGA

+1 R Y T P T S Q L R T Q E D A K E W L R S H S A G G L
81 CAGGTATACT CCCACCTCCC AGCTTCGCAC GCAAGAAGAT GCAAAAGAAT GGTACGGTC CCATTCTGCA GGAGGCCTTC
GTCCATATGA GGTGGAGGG TCGAAGCGTG CGTTCCTCTA CGTTTCTTCA CCAATGCCAG GGTAAGACGT CCTCCGGAAG

+1 Q D T A A N S P F S S G S S V T S P S G T R F N F S Q
161 AGGACACGCG TGCCAAATTC CCCTTTTCCT CTGGCTCCAG CGTGACTTCT CCCTCCGGAA CAAGATTCAA CTTTTCCTCAG
TCCTGTGCGC ACGGTTAAGG GGGAAAAGGA GACCGAGGTC GCACTGAAGA GGGAGGCCTT GTTCTAAGTT GAAAAGGGTC

+1 L A S P T T V T Q M S L S N P T M L R T H S L S N A D
241 CTTCGAGTTC CCACCACTGT CACCCAGATG AGCTTGTCAC ACCCGACCAT GCTGAGGACT CACAGCTCT CCAATGCTGA
GAACGCTCAG GGTGGTGACA GTGGGTCTAC TCGAACAGGT TGGGCTGGTA CGACTCCTGA GTGTCGGAGA GGTACGACT

+1 G Q Y D P Y T D S R F R N S S M S L D E K S R T M S
321 TGGGCAGTAT GATCCATACA CTGACAGCCG CTTCGCGAAT AGCTCCATGT CCCTGGATGA GAAGAGCAGA ACCATGAGCC
ACCCGTCATA CTAGGTATGT GACTGTCCGC GAAGGCCTTA TCGAGGTACA GGGACCTACT CTTCTCGTCT TGGTACTCGG

+1 R S G S F R D G F E E V H G S S L S L V S S T L S V Y
401 GTTCAGGCTC ATTCCGGGAT GGGTTTGAAG AAGTTCATGG ATCCTCACTC TCCTTGGTTT CCAGCACATT GTCAGTTTAT
CAAGTCCGAG TAAGGCCCTA CCCAACTTC TTCAAGTACC TAGGAGTGAG AGGGACCAAA GGTCGTGTAA CAGTCAAAATA

+1 S T P E E K C Q S E I R K L R R E L D A S Q E K V S A
481 TCTACACCAG AAGAAAAATG CCAGTCAGAG ATTTCGAAGC TGGCGCGGGA ACTGGATGCC TCCCAGGAGA AAGTTTCAGC
AGATGTGGTC TTCTTTTAC GGTCACTCTC TAAGCGTTTC ACGCCGCCCT TGACCTACGG AGGGTCTCT TTTCAAAGTCG

+1 L T T Q L T A N A H L V A A F E Q S L G N M T I R L
561 TTTGACCACC CAGCTGACAG CAAATGCTCA CCTTGTGGCT GCCTTTGAAC AGAGTCTTGG TAACATGACA ATCAGGCTCC
AAACTGGTGG GTCGACTGTC GTTACGAGT GGAACACCGA CGGAACTTG TCTCAGAAC ATTGTACTGT TAGTCCGAGG

+1 Q S L T M T A E Q K D S E L N E L R K T I E L L K K Q
641 AGAGTCTGAC CATGACAGCT GAGCAGAAGG ATTCAGAACT GAATGAGTTA AGAAAAACCA TTGAGCTGCT AAAGAAACAG
TCTCAGACTG GTACTGTCGA CTCGTCTTCC TAAGTCTTGA CTTACTCAAT TCTTTTGGT AACTCGACGA TTTCTTTGTC

+1 N A A A Q A A I N G V I N T P E L N C K G N G T A Q S
721 AACGCAGCTG CCCAGGCTGC CATTAATGGA GTAATTAACA CACCTGAGCT CAACTGCAAA GGAAACGGCA CTGCCCAGTC
TTGCGTCGAC GGGTCCGACG GTAATTACCT CATTAATTGT GTGACTCGA GTTGACGTTT CCTTTGCCGT GACGGGTCAG

+1 A D L R I R R Q H S S D S V S S I N S A T S H S S V
801 TGCAGACCTC CGCATCCGCA GGCAGCACTC CTCAGACAGC GTCTCCAGCA TCAACAGTGC CACCAGCCAC TCCAGTGTGG
ACGTCTGGAG GCGTAGGCGT CCGTCGTGAG GAGTCTGTCT CAGAGGTCGT AGTTGTACAG GTGGTCGGTG AGGTACACAC

+1 G S N I E S D S K K K K R K N W L R S S F K Q A F G K
881 GCAGCAACAT AGAGAGTGAC TCAAAGAAGA AGAAGAGGAA GAAGTGGTTA CGCAGCTCCT TCAAGCAAGC TTTCCGGAAG
CGTCGTTGTA TCTCTCACTG AGTTTCTTCT TCTTCTCTT CTTGACCAAT GCGTCGAGGA AGTTCTGTTC AAAGCCCTTC

+1 K K S P K S A S S H S D I E E T T D S S L P S S P K L
961 AAGAAGTCCC CAAAATCTGC GTCCTCTCAT TCAGATATTG AGGAGACGAC GGATTCTTCT TTGCCTTCTT CACCAAAGTT
TTCTTCAGGG GTTTTAGACG CAGGAGAGTA AGTCTATAAC TCCTCTGCTG CTAAGAAGA AACGGAAGGA GTGGTTTCAA

+1 P H N G S T G S T P L L R N S H S N S L I S E C M D
1041 ACCGCACAAAT GGGTCCACAG GTTCCACCCC ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTCCGAA TGCAATGGATA
TGCGGTGTTA CCCAGGTGTC CAAGGTGGGG TGACGACTCC TTAAGAGTGA GGTGAGAGA TTAAGGCTT ACGTACCTAT

+1 S E A E T V M Q L R N E L R D K E M K L T D I R
1121 GTGAAGCTGA GACCGTCATG CAGCTCCGAA ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CGGATATCCG C
CACTTCGACT CTGGCAGTAC GTCGAGGCTT TACTCAATTC TCTGTTCTC TACTTCGACT GCCTATAGGC G

11/24/97 02:46PM

hh2UNC53_5'Race_c13.7_CON.seq

Page 1

+2 E F E L G T L T I G L E R P P G Q V R D G F E E V H
1 CGAATTCGAG CTCGGTACAC TCACTATAGG GCTCGAGCGG CCGCCCGGGC AGGTCCGGGA TGGGTTTGAA GAAGTTCATG
GCTTAAGCTC GAGCCATGTG AGTGATATCC CGAGCTCGCC GCGGGGCCCG TCCAGGCCCT ACCCAAACCT CTCAAGTAC
.....
+2 G S S L S L V S S T S S V Y S T P E E K C Q S E I R K
81 GATCCTCACT CTCCTGGTT TCCAGCACAT CGTCAGTTTA TTCTACACCA GAAGAAAAAT GCCAGTCAGA GATTCGCAAG
CTAGGAGTGA GAGGAACCAA AGGTCGTGTA GCAGTCAAAT AAGATGTGGT CTTCTTTTA CGGTCACTCT CTAAGCGTTC
.....
+2 L R R E L D A S Q E K V S A L T T Q L T A N A H L V A
161 CTGCGGCGGG AACTGGATGC CTCACAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACA GCAAATGCTC ACCTTGTCG
GACGCCGCC TTGACCTACG GAGGTCCTC TTTCAAAGTC GAACTGGTG GGTGACTGT CGTTACGAG TGAACACCG
.....
+2 A F E Q S L G N M T I R L Q S L T M T A E Q K D S E
241 AGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG GACTCAGAAC
TCGGAAACTT GTCTCAGAAC CATTGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTG CACTGTCTTC CTGAGTCTTG
.....
+2 L N E L R K T I E L L K K Q N A A A Q A A I N G V I N
321 TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGCAGCT GCCCAGGCTG CCATTAATGG AGTAATTAAC
ACTTACTCAA TTCTTTTGG TAACTCGACG ATTTCTTGT CTTGCGTGA CGGGTCCGAC GGTAAATTACC TCATTAATTG
.....
+2 T P E L N C K G N G T A Q S A D L R I R R Q H S S D S
401 ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCAGT CTGCAGACCT CCGCATCCG AGGCAGCACT CCTCAGACAG
TGTGAGCTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GCGTAGGCG TCCGTCTGTA GGAGTCTGTC
.....
+2 V S S I N S A T S H S S V G S N I E S D S K K K K K R
481 CGTCTCCAGC ATCAACAGTG CCACCAAGCA CTCCAGCGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG AAGAAGCGGA
GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTGCGAC CCGTCGTTGT ATCTCTCACT GAGTTTCTTC TTCTTCGCCT
.....
+2 K N W V N E L R S S F K Q A F G K K K S P K S A S S H
561 AGAACTGGGT CAATGAGTTA CGCAGCTCCT TCAAGCAAGC TTTCGGGAAG AAGAAGTCCC CAAAATCTGC GTCCTCTCAT
TCTTGACCCA GTTACTCAAT CGGTCGAGGA AGTTCTGTCG AAAGCCCTTC TTCTTCAGGG GTTTTAGACG CAGGAGAGTA
.....
+2 S D I E E M T D S S L P S S P K L P H N G S T G S T P
641 TCAGATATTG AGGAGATGAC GGATTCTTCT TTGCCTTCT CACCAAAGTT ACCGCACAAT GGGTCCACAG GTTCCACCCC
AGTCTATAAC TCCTCTACTG CCTAAGAAGA AACGGAAGGA GTGGTTTCAA TGGCGTGTGA CCCAGGTGTC CAAGGTGGGG
.....
+2 L L R N S H S N S L I S E C M D S E A E T V M Q L R
721 ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTTCAGAA TGCATGGATA GTGAAGCTGA GACCGTCATG CAGCTCCGAA
TGACGACTCC TTAAGAGTGA GGTGAGAGA TTAAGTCTT ACGTACCTAT CACTTCGACT CTGGCAGTAC GTCGAGGCTT
.....
+2 N E L R D K E M K L T D
801 ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CCGATAT
TACTCAATTC TCTGTTCTC TACTTCGACT GCCTATA

11/24/97 03:29PM

E1.3-3

Page 1

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

11/24/97 03:34PM

hn25 raceB2-1.CON

Page 1

1 S M R G H T A O V G I T H P R T K A S A P A G A L
1 P U R E G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 T P G T O K T D P A K V S E G R L S P K A S O V K
81 G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 R S P S D A G R S S G D E S K K P L P S S S R T P T A
161 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 N A N S P G F K R Q S G S A A G L A M I T A S G V T
241 A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
T H A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 T S R S A T L G K I P R S S A L V S P S A G R K S S
321 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 M D G A O N O D O G V L A L S S R T N L Q Y R S L P R
401 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 P S R S N S R N G A G N R S S T S I D S N I S S R
481 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 A G L P V P K L R E P S K T A L G S S L P G L V N O
561 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 T D K E K G I S S D N E S V A S C N S V R V N P A A O
641 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 P V S S P A O T S L O P G A K Y P D V A S P T L R R
721 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 F G G K P T K O V P I A T A E N H K N S V I S N P
801 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 H A T M T O O G N L D S P S G S G V L S S G S S P L
881 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
T H A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 Y S K N V D L N O S P L A S S P S S A H S A P S N S
961 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 T M G T N A S S S S A V S R D G L G F O S V S G L H
1041 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 T S C E S I D I S L S S G V P S H N S S T G L I A S
1121 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
1 S K D D S L T P F V R T N S V K T T L S E S D P H L
1201 G E A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S

11/24/97 03:34PM

hn25 raceB2-1.CON

Page 2

1 R N T L P K K G L R V
1281 A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S
A G A R T A C T I O N S G E A R T A C T I O N S G E A R T A C T I O N S

Page 2

[illegible]

11/24/97 03:54PM

A2-2-2 (h2)NCS3 5 PAGE 02/04/97)

Page 1

*2

EcoRI
A R A G E S S S T P T

1 GTCGAAAT GCGTAACT ACTATGAG TCGAGGCG GCGCGGAG CTGAGAGG GAGAGTTC AGCTTAAAT

AGCTTTAA GCGAAATAG TGAATATTC AGCTGCGG GCGAGGCT GCTCTTCC GTCTGAGG TGGAAATAT

*2 N G S T S S A I P O P G A A T K P W R S K S L S V K H

81 ATTGAGAT GCTCTGCGG ATTGCGAG GCGGAGAG GAGAGTTC TGGAGAGG AATCTTCA GATAGAGC

TAGCTTAT GAGAGGCGG TGGAGGCTG GCGAGGCTG GTCTTCAAT ACCGCTGT TTAGGAGAT GAGTCTG

*2 S A T V S M L S V K P P G P E A P R P T P E A N K P A

161 AGTGGAGG TATCTATCT GTCGTGAG GTCTTCCG GAGAGTTC GAGTCTCA TTAGCTTC TTAGCTTC

TGAGCTTC ATAGATGAG GAGAGTTC GAGAGGCG GAGTCTTC GTGCGGAT GAGCTGAT ACTTGGCG

*2 P N N O R S H L E K L K L F N S K O G S K A G E G P

241 CCGCAAAAT GAGAGTCA TGTGAAAT GTGAAATC TTAGAGAT AAGAGGAT AAGAGGAT GAGAGGAT

GAGTCTTA GTCTGAGT AGAGCTTIT GAGCTTGA AGTCTTAT TTCTCTGAG TTCTCTGAG CTCTCTGAG

*2 G S R D T S C E R L E T L P S F E S E E L E A A S R

321 GATGCGGA GAGAGTGT GAGAGTTC AGAGTTCG GAGTCTCA GAGAGGAG AGTGGAGG GCGAGTTC

GAGAGGCT GTCTGAGC GTCTGAGC TCTGAGCG GTGAAATCT GTCTCTTC TCGAGTTC GCGAGTTC

*2 M L T T V G P A S S S P K I A L K G I A O R T F S R A

401 ATTCTACA GCGTGGCT TGTCTGAG AGCTTCCG AGCTTCCG GAGAGTTC GAGAGTTC TTAGCTTC

TAGAGTGT GCGAGGCGG AGAGAGTTC TCGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC

*2 L T N K K S S L K G N E K E F T Q O R E K F K

481 AGTGGAGG AGAGTTC GTCTGAGG GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC

TAGCTTC GTCTTCA GAGAGTTC GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA

*2 S R D L A K R A S V T E R L D L K E E P K E D P S G A

561 GAGAGTTC TGTGAAAT GTCTTCA GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC

GAGTTCAG AGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC GAGAGTTC

*2 A V P E M P K S S K I A S F I P K G G K L N S A R K

641 GGTGCGCG AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC

GAGAGTTC TTAGAGTTC TTAGAGTTC TTAGAGTTC TTAGAGTTC TTAGAGTTC TTAGAGTTC

*2 F P M A P S H S G I P R P G N F S M P G K S P S A P

721 GAGAGTTC GCGTTCG AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC AGAGTTC

GAGTTCAG GCGTTCG TGTGAAAT TGTGAAAT GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA

*2 A P S K E G E R S R S G K L S S G L P O Q K P O L D G

801 GGTCTTCA GAGAGGAG GCGTTCG GTGAAATC GAGTTCAG GTCTTCA GTCTTCA GTCTTCA

GCGTTCAG GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA

*2 R H S S S S S L A S S E G K G P G G T L N H S I S

881 AGAGTTC GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA

TGTGAAAT GAGAGGAG GTCTTCA AGAGTTC GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA

*2 T G E D P E A R R L R T V K N I A D L R O N I E E T
1201 GAGTTCAG GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA
GAGTTCAG GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA GTCTTCA
*2 H S S L R G T O V T K P N
EcoRI
1281 TGTGAGTTC AAGAGGAT GAGTTCAG AGAGTTC
AGAGTTCAG TTCTTCA GTCTTCA TGTGAAAT

11/24/97 03:54PM

A2-2-2 (h2)NCS3 5 PAGE 02/04/97)

Page 2

11/26/97 04:11PM D2.1-5 (hh2UNC5) 5'PAC 02/04/97) Page 4

1201 +3 G E D P E A R R L R T V K N I A D C R Q N L E E I V S
GCGAGATTC TAGATGCGC CCGCTCCGA CAGTACGAC CATCGTAT CTCGCCGAT ATTGACAGA MCGGTGTC
GCTTATGTC ATCGGACGCG CCGCATGCT GTCACTTCT GAGACGACT GAGCCGCTT TAAAGTCTT TTCGCACAG

.....
+J S L R G T Q V T K P N
ECORI

1281 AGTTTAAAGG GAACTCAAGT TACAAAGCG AATTC
TCMAATTCCT CTTCAGTCA AATGTTCCGC TTAAAG
.....

1201 +3 A P C Q P H Q P A P W O O S K A O A E M O S R I S G I K
GCTTCACGAG CAGACACAC ATACAGATC AMAGAGAA CTTAAATCC AGTCAGAT TCGAGTAC
GCGGACAGT CCGATGATC GATCTGATC TATCTGATC TTTCTGATC GATCTGATC TCGATCTC
+3 P N
1281 CCGAATTC
GCTTTAG

Fig. 11f.

1 GATATCTGCA GATATCTGCT TACTACTAT AGGCTGTCAG CCGATCTGTC GCGACATCTC TCGACATGAG GCTTAGAGAG
CTTAGAGCT CTTAAAGGCA ATAGATGATA TCGGATCTC GCGGACGCG CCGTCTGATC AGGCTCTGAG CAGCTCTC
+3 A A P R S R C G R Y G S Y T H K E D D S P G G E P
81 GCGCTCTCC TCGGCTGTC TATGGCTCTT ACACATGAA GAGATGATC AGCGAGAG AGGAGCAGC
CGCGAGGAG CTTGCTGAC ACCGCTCC ATACCGGAA TGTGTACTT CTTTACTG TCGGCTCTC CCGCTGTC
+3 L * T L A G A M R G G F P R S V C S D L W E S G A
161 GCTGATGTA AGGCTGTCAG GAGGATGAG GCGGCTCTT CCGGAGTC TTTGAGTGA TTTATGAGAG TCGAGGCTG
GAGCTACT TCGACATCTC CTTGCTACT CCGCGGAAA GCGGCTGAC AACCTGCT AACCTCTC AGTCTGAC
+3 A R E G C G C * Q L G G L N A H L H C Y * N I N Y L C
241 CCGAGAGAG ATGTGATGT TACAGCTG GTGATTTAA TCGGCTTTA CTTGTATTT GATATTTAA TTTCTGTC
GCTCTCTCC TACACTTACA ATGCTGAC GACTATTT AGGATTAAT GTAGATTTA CTTATTTT ATTAGACAG
+3 J Q G D S E G K G F W U S N V V C A S G Y G L V F A Q
J-1 CAAGAGAA GTAGAGAA GATTTTATC GAGATGATC TATGATGAT GACTATCTT TTTATGAT
GTATCTCT GATCTCTT CCGAAGAC CCGATTTAT ACATATGAG GATCTATA CTTAAAGCA AACGCTCTA
+3 U P L G S H E A S L K P K V K P Q D L A L M H P A A
481 TTTCTCTG AGCTCTAT AGGCTGCT CAGATCTAA GTAAACCC AGACTTAC TCTGATGTC CCGGACAG
ACGAGATC CCGAGATC TCTCTGAG GTTGTGAT CATTGTGAG TTTGATGAG AGATGATC GCGCTCTC
+3 V T L I A K A V P P S E V P E W I L R R Q E D L * R G
481 TACGCTGTA TCGAAGGCA GTGCTGCTC GTAGCTGCT CAGTGTAT TTAGAGAC AGGAGATCT TTAGAGCT
AGTGTGATG ACTTTGCT CAGCAGGCT CAGCTGAG GCTCAGTAA ATTTCTG TCTCTGCA ACTCTGCA
+3 G A D F L G W R E L C L A V A C I T F V N V I L F L
561 GCTCTGAT TCTCTGTC GCGGATC TCTCTGCT TCTGATCT CAGTGTAT TCGTCTCT TCTCTCTG
CAGCAGTAA AGGAGCAG CCGCTGAG AGAGAGCAG AGTATCTA GTAAAGCA ACCATGAC AGGAGAGC
+3 G F G S I A E G E R G P Y F W E H E S V S E S Q Q
641 TCGATTTGA AGCTGCTG AGGAGAG AGCTTTAT TCTGAGAA TCGATCTCT TCTGATCT AGCGAGAG
ACCTAACT TCTAGGAG TTTCTCTC TCTAAATA AGAGCTTT ACTTAGCA AGCTGAG TGTGTGTC
+3 O K R K P V I H G L E D Q K R I Y T D W A N H Y L A K
721 AGAGAGAG CCGATCTC CAGGATCTC AGATCTAA AGAGCTAC AGGATCTG CCAATCTA CTTAAAGCAA
TCTCTGCT CCGATGAG GTATCTGAG TCTGATTT CTTGATGTC TCTGATCTC GATGATCT GATGATCT
+3 S G H K R L I R D L Q Q D V T D G V L L A Q I I O V
801 TCGGCTGCA AGGCTGCT CAGGATCTC CAGGATCTC TCGAGATG TCGATCTCT GCGGATTA TCGAGTCT
AGGCTGCT TCGAGGTA GCTCTGAG GTGCTCTC ACTGCTAC CAGGATCTC CCGGATCTC AGTCTGCA
+3 A N E K I E D I N G C P K N R S Q M I F N I D A C L
881 CCGAATGAA AGATGATGAG AGATCTGAG CTTGCTGAG ACAGATCTC AMATGTA AMATGAT CCGTCTA
CGCTTACT TTTACTCT TATGATCT CAGAGCTCT TCTCTGAG TTTACTACT TTTGATCTA CCGAGATC
+3 N F L A A K G I N I O G L S A E E I R N G N L K A I L
961 ATTTCTGTC ACTTAGGTA ATAGATCT AGGCTGTC TCGAGAGAG ATGAGATG GAGATCTA GCGATCTA
TAGAGCTG TGTATCTCT TATTTGAG TCGGAGAG AGGCTCTC TATGCTCTC TTTGATCT CTTGATGAT
+3 G L F F S L S R Y K Q Q Q Q P Q K Q H L S S P L P
1041 GCGCTCTCT TACGCTCT CCGATGAG AGGAGAGAG AGGAGAGAG CAGAGAGAG CTTCTCTC CTTCTCTC
CGGAGAGAG AGTATGAT GATGATCT GATGATCT TCTGCTGCT TCTGCTGCT CTTCTCTC GAGAGAGAG
+3 A V S Q V A G A P S Q C Q A G T P Q Q V P V T P Q
1121 CCGCTATCT CAGTCTCTC GCGCTCTCT CAGTCTCTC CTTGAGAG CTTGAGAG CTTGAGAG ACTGCTG
CGGATGAG GTTCTCTCTC CCGGAGAG CCGGAGAG CAGGATCT CAGGATCT GATGATCT CAGGATCT TCGAGCTC

11/24/97 04:17PM

1121 AT:FC TAT: 42 142

1 CTTACAAAGH TTTTACGCA AGCGTTTAT GTCTCAGCC CGAGACAGG CTGTCTTGG AGHTGKCA GTTTCGATT
GAATTTTCH AACATCTGT TCCGMAA CAGATTCG GCTCTTCC GACAACTC TCAACGCT CGAAGGTAA
81 GCGCGGACA ACAGGCGAG TCGCTTACA GCAAGACAG NTGCTGGT TTTCTTGGC P P F P R D T G
CGGCGCTGT TGTCTGTC ACAGATCT CTTCTTGT MAGCACC AACAAGACG GCGAAGGTT CTTCTGTCC
3 P O T L F A E T S T A F I I T K P L P C S D N E F O
161 CAGAGAGT CTGTCTGT AGACTTCAC ACCTTTCAIT ATTACAAAC CTCTCCATG CTCAGACAT GAGTTTCAGC
CTCTCTGT GACAAAGAC TCTGAAGTG TCGAAGTAA TATGTTTG GAGAGGTAC GAGTCTGTA CTCGAAGTCG
3 O H A L L P R A S V E H L E * O L D F I E C L P C T
241 AATGTGCG ACTACTCG AGACATCTG TAGAGACT AGATACAA CTGACTTTA TTGAGTTT ACATGAC
TTATATCG TATGAAGC TCTGTAGAC ATCTCTGTA TTTATTTT GACTGAAT AATACAAA TATGAGTG
3 K P W A K H F I C R L F V F T A N P V G R Y N Y P H S
321 AGCTCTGG CTAAAGCT CATCTCAG CTGTGCTT TTACGCCAA CCGATAGT AGTATAGT ATCCGACT
TTGGAGCC GTATGTGA GTAGAGTCC GACAGAGA ATGCTGTT GGTATCCA TCGATATGA TAGGTTAG
3 A D A E T E A O S V L V A K O A H O E A R K W P H L
401 TAAATACA GAAAGGAG CACAGATGT TTGTATGT AACAAZT AACAAZC TAGAAGTG CCAAGCTAG
ACGTATAT CTCTTCTC GTCTTCTCA AACATGA TTCTTGAG TGTCTCTG ATCTTACC GGTGTGATC
3 A G P P D S T N C L P L L C C H O E C D S K F F L P S
461 CTGGCCCC TACTCTAC ACTGCTCG CTCTCTGT TCGATGAA GATGTGAT CCAATTTT CTTCTTCT
GACCGCGGG ACTGAGTG TTACGAGG GAAAGCAC AGCTAGCT CTATACUA GTTCAAAA GGAAGGAGA
3 G S N S G F T L L S N O I Y T D W A N H Y L A K S G H
561 GATTCAGT CTGCTCAC TGTCTGAC ACCAGATG GGCATGCT TACTAGCA AATCCGCA
CTTAGTTA GACGAATG AAGAGTGT TTGTCTGA TGTCTCTG CCGTTGTA ATGATGCT TTAGCGCT
3 K R L I K D L O O E S R I P A H W R F L L V D P S S
641 CAGCTATC ATCAAGAT TCGAGATA AGCGGATG GCGAGACT GCGCGCTT ACTAGCAT CCAAGCTCG
GTTGAGAG TATGCTAG AGTGTGCT TTGGTTAA GGTCTGTGA CCGCGGGA TATGACTA GCTTCAGCC
3 V P K L G V I M V I V
721 TACCAAGT TGGCTATC ATGCTATG TGT
ATGCTTGA ATCGATAG TACGATAC ACA

11/25/97 01:48PM

hrzunc_f4_1_2.m13 Seq

Page 1

1 NHAAGAGA GGGGTAAA AAGGTATTA AHAANAG GATCGTTTG NGGGGTATTT TTTTNGCC CCGGGTTC
NCTTCTCT CCGCATTTT TTNCAJAN TTTTNTTTC GTAGGAAAC NGCCCAAA AAAANKGG GGGNCAAC
81 TAAAGGGG TCGGTAAAG ATTTTAAAG GAGGGAGG GAGGTGCT TCGGTGCT TCGGTGCT GTCANAGG
ATTTTCCCC AGGGATNCC TAAAGATNCC GTTCTGCT CTTCTGCT TCGAAGAA ACCGNCNCC CACTTGGC
161 NCACACCT GGGGAGGAG NGGAGCAAC TTTTCTTTT AHAAGAGA ATTTTGAAG AAGGGGNG TTTGTAAG
NGTGTGGA CCGGCTGCT NGGCTGCTN AHAAGAAA TTTTCTCT TAACTTCT TTTTCTCT AAGCTATC
241 NTGGGAGT TTTGGGNC AHAAGATG AGAGTANG TGGATGCG GGTAGGNG GGGATATC NGGGTCTT
HAACCTCA AAGGCGNG TTAATTTAC TGTCTATCC AGCTATGCG GGTATGCG GGTATGAG NGGCAAGA
321 AAGTNGGG TTTGTTTC CAGGTGCA AGGTGAGG TTAAGAA TCCATATA CTGGTTAA TGAATCTT
TCAATCCC AHAAGAGG GTNAGCTT TTTCTTTC AATTTCTTA AGGTATTT GAGGATTT ACTTGTAA
401 TGAATTTT TTAATAAA ATAAATGAA TAAATGAA TTTCTTNC CAGAAAAA AHAAGAGA AHAATACA
ACTAAAAA AHAATATTT TATTAATTT AATTTTAA AHAAGAGG GTTTTTTT TTTTCTCT TTAATTTT
481 HAATANTA HAHAHAAN GAHAAGAN HAATAAAA HAHAAGAG GGGGTAAAT TTAAGCTT TAAAGATG
NTTAATTA TTTTATTA CTTTCTTA TTTTCTTT TTTTCTCT GGGGCTTA AHAAGAGG ATCTGTC
561 UBP HSL S BDR VVR VPA PIV S * Y F G V
TGTGCTCT CAGGCTCT CCGGAGAG AGGTGCTG GTCCCGGAG CCGGCTAT GTCTGATC TTTGGGCT
ACGCGGA GTTGAGCA GGGGCTGTC TACCAAGCG CAGGGGCTC GGGGTACA CAGACTAG AAGCGGAG
641 HMAIDL YCGLA CLW GKH EPR IYTD MAN
AGATGCTAT AGATTTTC TCGGTTCG GTTCTCTC GGAAGATC GAGCGGAA TCTACAGA CTGGGCAAT
TGTACCTA TTAAGATG AGGAGCC GAGGAGAC CCGTTCTA CTGGGCTT AGATGCTC GAGCGGTA
721 HYLAKSGHKRIKDFQESRLQISIT
CATACTAG CAAATGCG CAAAGCTG ATCTACAG ATTTGAGA HAAGAGA ATCTGAGA TATCATAC
GTATGAGC GTTTAGGC GTTTGAGA TATGATTC TAAAGTCT TTTTCTCT TTAAGTCT ATAGTATG
801 XAARACX * RPIRPIVS C I Q
ATGCGGCC CCGTGAAG GATTTAGG CCGATGCG CTTATATG GTTATACA AT
TAAAGCGG CAGCTCTA CTAAATCT CCGTAAAG GATATGCT CAGATATG TA

Figure 12. Tblastn search of the EST division of Genbank with 680aa starting at the c-terminus of the alfa-actinin domain of Hu-UNC-53/2.

LOCUS AA418158 610 bp mRNA EST 19-MAY-1997
 DEFINITION zv97d12.r1 Soares NhHMPu S1 Homo sapiens cDNA clone 767735 5'.
 ACCESSION AA418158
 NID g2079968
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 610)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
 Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLE WashU-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 492.
 FEATURES
 source Location/Qualifiers
 1..610
 /organism="Homo sapiens"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NbHM, pregnant uterus
 NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 /clone="767735"
 /clone_lib="Soares NhHMPu S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 mRNA <1..>610

```

                                /clone="5D16"
                                /clone_lib="Zebrafish ICRFzfls"
                                /sex="mixed"
                                /tissue_type="pooled 26-somite embryos"
                                /lab_host="XL1-blue MRF"
                                complement(<1..>418)
mRNA
BASE COUNT      108 a      87 c      78 g      145 t
ORIGIN
      1 tttacatttt ttgaggaaga tgctaattggt ctattctgat tcaatgattt atgctaagct
     61 aagctaaaaat gctcctgtca aatcctgaga tcagctgaat gaattaaaaa tttggtaaaa
    121 ctcaactgtc taactctagg ggagttgtaa aatgggccta ttccctaaa aagtaatggt
    181 actttaagag catgatgggc caccagtttc actgtctaaa tttgttatt ccataagcta
    241 atcttctctg ggcattttga cgattttaac actaacctgt gggtaatctg cgtccccgt
    301 aaactggaca tggttcttc cagattctgt ctcagatcag caatgttctt cactgtacgc
    361 atccgtctag tttctggatc ttctcctgag atctcctcca ggcactgttt ggcggtct

//
gb|AA495042|AA495042 fa05f06.s1 Zebrafish ICRFzfls Danio rerio cDNA
      clone 5D16 3'
      Length = 418

```

Minus Strand HSPs:

Score = 195 (87.9 bits), Expect = 9.9e-18, P = 9.9e-18
 Identities = 37/46 (80%), Positives = 42/46 (91%), Frame = -3

Query: 627 TGQPALEELTGEDPEARLRVKNIAADLRQNL EETMSSLRGQTQVTH 672
 T + LEE++GEDPE RR+RTVKNIAADLRQNL EETMSSLRGQTQ+TH
 Sbjct: 416 TAKQCLEEISGEDPETRRMRTVKNIAADLRQNL EETMSSLRGQTQITH 279

MOUSE 2

LOCUS AA208994 527 bp mRNA EST 18-FEB-1997
 DEFINITION mw75e12.r1 Soares mouse NML Mus musculus cDNA clone 676558 5'.
 ACCESSION AA208994
 NID g1807004
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.
 REFERENCE 1 (bases 1 to 527)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:416262

Putative full length read

vector to vector length is 535

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 478.

FEATURES

source

Location/Qualifiers

1..527

/organism="Mus musculus"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCGAATCTTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

/clone="676558"

/clone_lib="Soares mouse NML"

/tissue_type="Liver"

/lab_host="DB10B"

<1..>527

mRNA

BASE COUNT

151 a 139 c 136 g 101 t

ORIGIN

```

1  tgtctctgga  tgagaagagc  cgaacaatga  gtcgggtcagg  ctccttccgg  gatggggttg
61  aggaagttca  tggatcctcc  ctgtccttgg  tttccagcac  atcctccatc  tactccacgc
121  cagaagaaaa  atgccagtca  gagattcgaa  agctgaggcg  agacgtggat  gcctcccagg
181  aaaagggtgc  tgcgctgact  acccagctga  ctgcaaatgc  tcaccttggtg  gcagccttcg
241  agcagagtct  gggaaacatg  accatcaggc  tacagagttt  aactatgacc  gctgagcaga
301  aggattcaga  actgaacgag  ttaagaaaaa  ccatcgagct  gctgaagaaa  cagaatgcag
361  ctgcccaggc  tgccattaat  ggagtgatta  acacgccaga  gctcaactgc  aaaggaaatg
421  gcagtgccag  gctacagacc  tacgcatccg  cagcaacact  cctccgacag  tgtctccagt
481  atcaatagcg  ccaccagcca  ctcaagtgtg  ggcagcaaca  tagagag

```

gb|AA208994|AA208994 mw75e12.r1 Soares mouse NML Mus musculus cDNA
clone 676558 5'
Length = 527

Plus Strand HSPs:

Score = 541 (243.9 bits), Expect = 2.3e-76, Sum P(2) = 2.3e-76
Identities = 110/143 (76%), Positives = 114/143 (79%), Frame = +3

Query: 1511 SLDEKSRTMSRSGSFRDGFEEVHGXXXXXXXXXXXXXXXXXXXXPEEKQSEIRKLRRRLDASQE 1570
SLDEKSRTMSRSGSFRDGFEEVHG PEEKQSEIRKLRR++DASQE
Sbjct: 3 SLDEKSRTMSRSGSFRDGFEEVHGSSLSLVSTSSSIYSTPEEKQSEIRKLRRDVSASQE 182
Query: 1571 KVSALTQTQTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDSSELNLRKTIEXXXXXXXX 1630

Sbjct: 183 KVSALTTQLTANABLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIE
KVSALTTQLTANABLVAAFEQSLGNMTIRLQSLTMTAEQKDSELNELRKTIELLKKQNAA 362

Query: 1631 XXXXXXGVINTPELNCKGNGTAQ 1653
GVINTPELNCKGNG+A+

Sbjct: 363 AQAAINGVINTPELNCKGNGSAR 431

Score = 116 (52.3 bits), Expect = 2.3e-76, Sum P(2) = 2.3e-76
Identities = 24/25 (96%), Positives = 25/25 (100%), Frame = +1

Query: 1661 RQHSSDSVSSINSATSHSSVGSNIE 1685
+QHSSDSVSSINSATSHSSVGSNIE

Sbjct: 451 QQHSSDSVSSINSATSHSSVGSNIE 525

LOCUS AA049124 337 bp mRNA EST 09-SEP-1996
DEFINITION mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone 479167 5'.
ACCESSION AA049124
NID g1528794
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-BHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-BHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:289911
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 292.
FEATURES
source Location/Qualifiers
1..337
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGGCCGCGAAATTTTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

/clone="479167"

/clone_lib="Soares mouse embryo NbME13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

<1..>337

mRNA
BASE COUNT 80 a 101 c 97 g 59 t
ORIGIN

```

1 catcctctgt gggcaccgag gtcaccgaga cccctgctca ttcagtcctc cacactagac
61 tgttccaagc caatgaagag gaggagccag agaagaagga ggtatcagaa ctgcgctctg
121 aactatggga aaaagagatg aagctcacgg atatccggtt ggaggccctc aactctgccc
181 accagctgga ccagcttcgg gagaccatgc acaatatgca gttggagggtg gacctgctga
241 aagcagagaa tgaccggctg aaggttgccc ccgggccctc ctcaggctgc actccagggc
301 aggtccctgg gtcacggct ctgctgctcc ctgcgacg

```

gb|AA049124|AA049124 mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus
musculus cDNA clone 479167 5'
Length = 337

Plus Strand BSPs:

Score = 206 (92.9 bits), Expect = 3.9e-19, P = 3.9e-19
Identities = 42/60 (70%), Positives = 51/60 (85%), Frame = +3

Query: 1760 DSEAETVMQLRNELRDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLK 1819
+ E + V +LR+EL +KEMKLTDIRLEAL+SAHQLDQLRE M+ MQ E++ LKAENDRLK
Sbjct: 84 EPEKKEVSELSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLKAENDRLK 263
//

LOCUS AA185349 348 bp mRNA EST 07-JAN-1997
DEFINITION mu51c03.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
642916 5'.
ACCESSION AA185349
NID g1769059
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 348)

AUTHORS Marra,M., Billier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-BHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
 WashU-BHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:394908
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 336.

FEATURES Location/Qualifiers

source 1..348

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATACTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

/clone="642916"

/clone_lib="Soares mouse lymph node NbMLN"

/sex="male"

/dev_stage="4 weeks"

/lab_host="DH10B"

mRNA <1..>348

BASE COUNT 93 a 95 c 78 g 82 t

ORIGIN

1 attcggcact gaggggatga ataatccacc aaattagtgt gtacatagga gttgctgggc

61 cccccccac tttatctgc ttagctagc ctctccctaa gcctcgcatc ttctctaaat

121 ctatctctgc gttcttacca cttgttctgg ccaatagaac tccggatcaa gaggcagaat

181 tcctcagata gcatctccag cctcaacagc atcaccagcc attccagcat cggcagcagc

241 aaagatgctg atgccaagaa gaaaaagaag aagagttggg taagtaaagg cttggagata

301 ggcctgtgct aggagtcact caccctgttg cagggaactg accccctt

//

gb|AA185349|AA185349 mu51c03.r1 Soares mouse lymph node NbMLN Mus
 musculus cDNA clone 642916 5'
 Length = 348

Plus Strand HSPs:

Score = 154 (69.4 bits), Expect = $4.4e-12$, P = $4.4e-12$
Identities = 27/42 (64%), Positives = 40/42 (95%), Frame = +1

Query: 1656 DLRIRRQHSSDSVSSINSATSHSSVGSNIESDSKRRKNWL 1697
 +LRI+RQ+SSDS+SS+NS TSHSS+GS+ ++D+KKKK+K+W+
Sbjct: 157 ELRIKRONSSDSISSLNSITSHSSIGSSKDADAKKKKKSWV 282

FIGURE 12a

"SIM output with parameters:
substitution scores in BLOSUM62
O = 12, E = 4"

Sequence 1: hu1, 1702 residues
Sequence 2: hu2, 2350 residues

List of local alignments with score >= 100.0

46.8% identity in 1726 residues overlap; Score: 2538.0; Gap frequency: 9.3%

```
hu1,      78 DPESQRKRTVQNVLDLRQNLLEETMSSLRGSQVTHSSLEMTCYDS--DDANPRSVSSLSNR
hu2,     639 DPEARRLRTVKNIADLRQNLLEETMSSLRGTQVTHSTLETTFDTNVTTEMSSGRSILSLTGR
          *** * *** * ***** * * * * * * * * * * * * * * * * * * * * *
hu1,     136 SSPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMRSPSKLSHIS
hu2,     699 PTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLASRGSS
          ***** * * * * * * * * * * * * * * * * * * * * *
hu1,     196 RLEL-VESLDSDEVDLKS-----GYMSDSLGMKMTEDDDITTG-----
hu2,     759 VCHVDVSDKAGDEMDLEGISMDAPGYMSDGDVLSKNI-RTDDITSGYMTDGGGLGLYTRRL
          * * * * * * * * * * * * * * * * * * * * *
hu1,     235 -----WDESSSISSGLSDASDNLSSEEFNASSSLNLSLP
hu2,     818 NRLPDGMAVVRETQLQRNTSLGLGDADSWDDSSSVSSGISDTIDNLSDDINTSSSISSYA
          * * * * * * * * * * * * * * * * * * * * *
hu1,     268 STPTASRRNSTIVLRTDSEKRS LAESGLSWFSESEEKAPKKLEYDSGLKMEPGTSKWRR
hu2,     878 NTPASSRKNLDV--QTDAEKHSQVERNLSWGGDDVKKSDGGS--DSG-IKMEPG-SKWRR
          ** * * * * * * * * * * * * * * * * * * * * *
hu1,     328 ERPESCDDSSKGGELKKPISLGHGPGSLKKGKTPPVAVTSPITHTAQ--SALKVAGK---P
hu2,     932 NPSDVSDSDSKSTSGKKNPVISQTGSWRRGMTAQVGITMPRTKPSAPAGALKTPGTGKTQ
          * * * * * * * * * * * * * * * * * * * * *
hu1,     383 EGKATDKGKLAVKNTGLQRSSSDAGRDRLSDAKKPPSGIARPSTSG--SFGYKKPP-PAT
hu2,     992 DAKVSEKGR LSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPTANANSFGFKKQSGSAA
          * * * * * * * * * * * * * * * * * * * * *
hu1,     440 GTATVMQTG-----GSATLSKIQKSSGIPVKPVNGRKTSLDVSNNAEPGFLAPGARSNIQ
hu2,    1052 GLAMITASGVTVTSRSATLGKIPKSSAL-VRSAGRKSSMDGAQNQDDGYLALSSRTNLQ
          * * * * * * * * * * * * * * * * * * * * *
hu1,     495 YRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSSLSTKQGGLTPSRLKEPTKVASGRTPA
hu2,    1111 YRSLPRPSKSNRNGAGNRSS-----TSSID-SNISKKSAGLPVKLREPSKTALGSSSLPG
          ***** * * * * * * * * * * * * * * * * * * * * *
hu1,     555 PVNQTDREKEKAKAKAVALDSDNISLSKISGSPSTPKNQASHPTATKLAELPPTPLRATA
hu2,    1166 LVNQTDKEKGISSDNESVASCNSVKVNPAAPVSSPAQTS LQPGAKYPDVASPTLRLRFLG
          ***** * * * * * * * * * * * * * * * * * * * * *
```


hu1, 615 KSFVKPPSLANLDKV-NSNSLDLPSSSDTTHAS--KVPDLBATSSASGGPL-----P
hu2, 1226 GKPTRQVPIATAENMKNSVVISNPHATMTQQGNLDSPSGSGVLSSGSSSPLYSKNVDLNO
* * * * *
hu1, 664 SCFTSPAPILNINSASFQGLELMGFSVPKETRMYPKLSGLHRSMESLQMPMS---LP
hu2, 1286 SPLASSPSSAHSAPNSLWTGNTASSSSAVSKDGLGFQSVSSLTSCESIDISLSSGGVP
* * * * *
hu1, 721 SAFPSSPTVPTPPAPPAAP-TEEETEELTWSGSPRAGQLDSNO-----RD
hu2, 1346 SHNSSTGLIASSKDDSLTPFVRTNSVKTTLSSESPLSPAASPKFCRSTLPRKQSDPHLD
* * * * *
hu1, 765 RNTLPKKGLRY----QLQSQEETKERRHSHTIGGLPESDDQSELPSPPALPMSLSAKGQL
hu2, 1406 RNTLPKKGLRYTPTSQLRTQDAKEWLRSHSAGGLQDTAANSPPSSGSSSVTSPSGTRFNF

hu1, 821 TNIVSPTAAT-----TPRITRSNSIPTHEAAFEYSGSQM-GSTLSLAERPCKMIRSGSF
hu2, 1466 SQLASPTTVTQMSLSNPTMLRTHSLSNADGQYDPYTDSRFRNSSMSLDEKSRMTMSRSGSF
* * * * *
hu1, 875 RDPTDDVBGSVLSLASSASSTYSSAEERMQSEQIRKLRLRELESSQEKVATLTSQLSANAN
hu2, 1526 RDGFEEVHGSSSLVSVSTLSVYSTPEEKQSE-IRKLRLRELDASQEKVSALTQTQTANAE
* * * * *
hu1, 935 LVAAFEQSLVNMTSRLRHLAETAEEKDTLLELDRETIDFLKKKNSEAAVIOGALNASET
hu2, 1585 LVAAFEQSLGNMTIRLOSLTMTAEQKDELSNELRKTIELKKQNAQAQAINGVINTPEL

hu1, 995 TPK-----ELRIKRONSSDISLSNSITSHSSIGSSKADAKKKKKKSWVYELRSSF
hu2, 1645 NCKGNGTAQSADLRIRRHSSDSVSSINSATSHSSVGSNIESDSKKKKRKNW---LRSSF
* * * * *
hu1, 1047 NKAFSIKKGPKSASSYSIDIEEATPDSSAPSSPKLQBGSTETASPSIKSSTLSSVGTDTV
hu2, 1702 KQAFGKKKSPKSASSHSDIEE--TTDSSLPSSPKLPHNGSTGSTPLLRNHSNSL-----
* * * * *
hu1, 1107 EGPAHPAPHTRLFHANEEEEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAHQLDQLRET
hu2, 1755 -----ISECMDSEAEVMTQLRNELRDKEMKLTDIRLEALSSAHQLDQLREA
* * * * *
hu1, 1167 MHNMQLEVLLKAENDRLKVAPGPSSGSTPGQVPGSSALS-SPRSLGLALTHSFGPSLA
hu2, 1801 MNRMQSEIEKLAENDRLK---SESQSGGCSRAPSQVSIASPRQSMGLS-QHSLNLTES
* * * * *
hu1, 1226 DTDLPMDGISTCGPKKEVT--LRVVVRMPPQHIKGLKQOEFLGCSKVSGKVDWKML
hu2, 1857 TSLDMLLDDTGECSARKEGGRHVKIIVVSFQEEMKWKEDSRPHFLIGCIGVSGKTKWDVL
* * * * *
hu1, 1284 DEAVFQVFKDYISKMDPASTLGLSTESIHGYSISHVKRVLDAEPPEMPPCRRGVNN---I
hu2, 1917 DGVVRRLFKEYIIHVDPVSQLGLNSDSVLGYSIGEIKRSNTSETPELLPCGYLVGENTTI
* * * * *

```
hu1,      1341 SVSLKGLKEKCVDSLVEFETLIPKPMQHYISLLKHRRVLVSGPSGTGKTYLTNRLEAYL
hu2,      1977 SVTVKGLAENSLDSLVEFSLIPKPIQRYVSLLEHRIILSGPSGTGKTYLANRLSEYI
          ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1401 VERSGREVTEGIVSTFNMHQSCDKLQYLSNLANQIDRETGIGDVPLVILLDDLSEAGS
hu2,      2037 VLREGRELTGVIATFNVDHKSSKELRQYLSNLADQCNSENNAVDMLPVIILDNLHVSS
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1461 ISELVNGALTCKYHKCPYIIIGTTNQPVKMPNHEGLHLSFRMLTFSNNVEPANGFLVRYLR
hu2,      2097 LGEIFNGLLNCKYHKCPYIIIGTMNQATSSSTPNQLBHNFRWVLCANHTPEVKGFLGRFLR
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1521 RKLVESDS DINANKEELLRVLDWVPKLVYHLBTFLFKHSTSDFLIGPCFFLSCPIGIEDF
hu2,      2157 RKLMEITEISGRVRNMEVLKIIDWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGS
          *** * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1581 RTWFIDLWNNIIIPYLOEGAKDGKIVHGQKAAWEDPVEWVRDTLPWPSAQDQDS--KLYH
hu2,      2217 RVWFTDLWNYSIIPYLLEAVREGLQLYGRRAPWEDPAKVVMDTYPWAASPOQHEWPPLLO
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1639 LPPPTVGPBSIASPPEDRTVKDSTPSSLDSDPLMAMLLKLQEAANY
hu2,      2277 LRPEDVGF DGYSMPREGSTSKQMPPSDAEGDPLMNMMLRLQEAANY
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

WARNING: 49 local alignments have not been reported because of score < 100.0

"SIM output with parameters:
substitution scores in BLOSUM62
O = 12, E = 4"

Sequence 1: Cel, 1583 residues
Sequence 2: hu2, 2350 residues

List of local alignments with score >= 54.0

32.8% identity in 504 residues overlap; Score: 490.0; Gap frequency: 6.9%

```
Cel,      1058 VIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLKTENKQLKKEVDKLTNGPATR
hu2,      1766 VMQLRNELRDKEKMLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLAENDRLKSESQGS
          * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
Cel,      1118 ASSRASIPVIYD-----DEHVDYDAACSST-----SASQSSKRSSGCSNIKVTNV
hu2,      1826 GCSRAPSQVSISASPRQSMGLSQHSLNLTSTSLDMLLDDTGECSARKEGGRHVKIVVSF
          *** * * * * * * * * * * * * * * * * * * * * * * * * * * *
Cel,      1163 DIAGEISSIVNPDKEIIVGYLAMSTSQSCWKDIDVSILGLFEVYLSRIDVEHQLGIDARD
```

hu2, 1886 QEEMKWKEDSRPHL-FLIGCIGVS-GKTKWDVLDGVVRRLFKEYIIHVDPVSQLGLNS-D
* * * * *
Cel, 1223 SILGYQIGELRRVIGDSTTMITSHPTDILTSSTTIRMFMEGAAQSRVDSLVDMLLPKQM
hu2, 1943 SVLGYSIGEIKRSNTSETPELLPCGY-LVGENTTISVTVKGLAENSLDSLVSFESLIPKPI
* * * * *
Cel, 1283 ILQLVKSILTERRLVLGATGIGKSKLAKTLAAYVSIRTNQ--SEDSIVNISIPENNKEE
hu2, 2002 LQRYVSLLEHRRRIILSGPSGTGKTYLANRLSEYIVLREGRELTGVIATFNVDBKSSKE
* * * * *
Cel, 1341 LLQVERRLEKILRSKESCI-----VILDNIPKNRIAFVVSVFANV-PLQNNEGPFVVCVT
hu2, 2062 LRQYLSNLADQCNSENNAVDMPVLVIIDNL--HBVSSLGEIFNGLLNCKYHKCPYIIGTM
* * * * *
Cel, 1395 NRY--QIPELQIHNFKMSVMSNRLE---GFILRYLRRRAVEDEYRLTVQMPSELFKIID
hu2, 2120 NQATSSTPNLQLHNFVRWLCANHTEPVKGFGLGRFLRRKLMETEISGRVRN-MELVKIID
* * * * *
Cel, 1450 FFPIALQAVNNFIEKTSVSDVTGPRACLNCLPTVDGSRWFIRLWNNFIPYLERVARD
hu2, 2179 WIPKVWHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVRE
* * * * *
Cel, 1510 GKKTFGRCTSFEDPTDIVSEKWPW
hu2, 2239 GLQLYGRRAPWEDPAKWVMDTYPW
* * * * *

35.5% identity in 112 residues overlap; Score: 165.0; Gap frequency: 1.8%

Cel, 11 IYTDWANRHLSKGSLSKSIRDISNDFRDYRLVSQLINVIVPINEFSPAFTKRLAKITSNL
hu2, 11 IYTDWANHYLTKSGHKRLIKDLQDDVTDGVLLAQIIQVVA--NEKIEDINGCPKNRSQMI
***** * * * *
Cel, 71 DGLETCLDYLNKGLDCSKLTKTDIDSGNLGAVLQLLFLSTYKQKLRQLKK
hu2, 69 ENIDACLNFLAAGINIQGLSAEEIRNGNLKAILGLFFSLSRKQQQQQPOK
* * * * *

24.8% identity in 163 residues overlap; Score: 80.0; Gap frequency: 3.7%

Cel, 877 GSQSLASTT--AYGSLNEKYEHAIKRDMDLECYKNTVDSLTKKQENYGALFDLFEQKL
hu2, 1534 GSSLVSVSTLSVYSTPEEKQSEIRKLRLRELDASQEKVSALTQTQTANAHLVAAFEQSL
* * * * *
Cel, 935 RKLTOHIDRSNLKPEEAIKFRQDIABLRDISNHLASNSAHANEGAGELLRQPSLESVASH
hu2, 1594 GNMTIRLQSLTMTAEQK---DSELNELRKTIELKKQNAQAINGVINTPELNCKNG
* * * * *
Cel, 995 RSSMSSSSKSSKQEKISLSSFGK-NKKSIRSSLSKFTKKKNK
hu2, 1651 TAQSADLRIRRHSSDSVSSINSATSHSSVGSNIESDSKKKKR
* * * * *

58.6% identity in 31 residues overlap; Score: 74.0; Gap frequency: 6.5%

F

```
Cel,      653 GYPDNFEDSSSLSSGISDNNELDDISTDDL  
hu2,      840 GDADSWDDSSSVSSGISDT--IDNLSTDDIN  
          * *      * * * * *      * * * *
```

42.9% identity in 60 residues overlap; Score: 64.0; Gap frequency: 6.7%

```
Cel,      984 RQPSLESVASHRSSMSSSSKSSKQEKISLSSFGKNKSWIRSSLK-FTRKKKNKNYDEAH  
hu2,      1661 RQSSSDSVSSINSATSHSSVGS---NIESDSKXKKRKNWLRSSFKQAFGKKKSPKSASSE  
          ** * ** * * * * *      * * * * *      * * * *
```

22.0% identity in 91 residues overlap; Score: 56.0; Gap frequency: 0.0%

```
Cel,      140 SKLPSPRVATSATASATNPNNSNFPQMSTSLQTPQSRISKIDSSKIGIKPKTSGLKPPSS  
hu2,      177 SRLSGPTARVSAAGSEAKTRGGSTTANNRRSQSFNNYDKSKPVTSPPPPPSSEEKEPLAS  
          * * *      * *      * *      * *      * *      * *      * *      * *
```

```
Cel,      200 STTSSNNTNSFRPSSRSSGNNNVGSTISTSA  
hu2,      237 SASSEPGMSDNAPASLESGSSSTPTNCSTSS  
          * *      * * * * *      * * * *
```

WARNING: 44 local alignments have not been reported because of score < 54.0

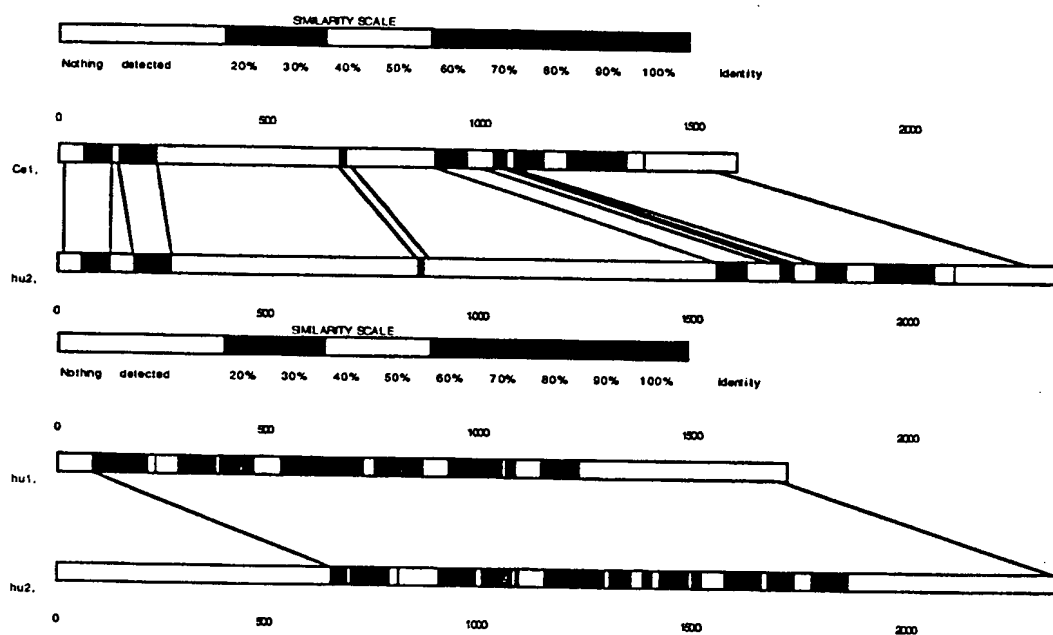


Figure 12b

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence

Enzymes : 100 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page

fig 13 pCB201

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGC GCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTGTGG
CCTCCAGCGACTCATCAGCGCTCGTTTAAATTCGATGTTGTTCCGTTCGAACTGGCTGTAACTGACTTCTTAGACGAATCCCAATCCGCAAAACGG
G G R V V R E Q N L S Y N K A R L D R Q L H E E S A G A F C

CTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGCTCTATATGCGCAACTGTAACATAAATGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
A A S R C T G Q I Y A L T L I I D L L I V I N Y G V I S S P I Y

TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGCGGGTTGCTGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K W P A W L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGACTATTTACGGTAACTGCCCACTTGCGAGTACATCAAGTGTATCATATGCCAAGTACGCC
TTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAAATGCCATTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGSTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTCTTACTTGGCAGTACATCTACGTATTAGTCA
GGATACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT
P Y R Q R M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGATGCGSTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCACATAGCCAAAACCGTCATGTAGTACCCGACCTATCGCCAACTGAGTGCCCTTAAAGGTTACAGAGGTGGGGTAAGTGCAGT
P Y Y H G D A V L A V H Q W A W I A V L T G I S K S P P H R Q

TGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCAAAATGTGTAACAACTCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGTTGGGAG
ACCTCAAAACAAACCGTGSTTTTASGTTGCCCTSAAGGTTTACAGCATGTTGAGGCGGGTAAGTGCCTTTACCCGECATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S Q L R P I D A N G R A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGASAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCCTATAGGGAGACCCAAAGCTGGCTAGC
CAGATATATCTGCTCGAGAGACCGATTGATCTCTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGACACCGATCG
G L Y K Q S S L A N R T H C L L A Y R N Y D S L G D P S W L A

GTTTAAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAATGGGTGGGATCTGTACGAG
CAAAATTTGAATTCGAATGGTACCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACTGTCGTTTACCCAGCCCTAGACATGCTG
F K L K L T M G G S H H H H H G M A S M T G G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GATGACGATAAGGTACCTAGGATCCATATGCCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGCTGAAGGAGAAATGCGTCGACAA
CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTCCAGACTTCCTCTTTACGCAGCTGT

pCB201 insert = U4

U4 ORF

D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C V D

GCCTGGTGTTCGAGACGCTGATCCCAAGCGGATGATGCAGCAC TACATAAGCCTCCTGCTGAAGCACC GGCGCTCGTCTCTCGGGCCCCAGCGGCAC
CGGACCACAAGCTCTGCGACTAGGGGTTTCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCCGGGTGCGCGTG

pCB201 insert = U4

U4 ORF

S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S G P S G T

GGGCAAGACCTACCTGACCAATCGCTTGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCA
CCCCTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGT

pCB201 insert = U4

U4 ORF

G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q

CAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACAGATAGACCGGAAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGAGATG
GTCAGAACGTTCTAGACGTTGACATAGAAAGGTTGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTAC

pCB201 insert = U4

U4 ORF

D S C K D L O L Y L S N L A N O I D R E T G I S D V P L V I L L D

AGCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCTATATTATAGGTACCAACCAATCAGCTGT
TGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGAGTGGACGTTTCATAGTATTACAGGSATATAATATCCATGGTGGTTAGTCGGACA

pCB201 insert = U4

U4 ORF

D S E A G S I S E L V N G A L T C K Y H K C P Y I I G T T N Q P V

AAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGSATGTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCTGGTTGCTTACCTGAGG
TTTTTACTGTGGGTTGGTACCGAACGTGAACCTCAAGTCTACAAC TGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGAATCC

pCB201 insert = U4

U4 ORF

K M T P N H G L H L S F R M L T F S N N V E P A N G F L V R Y L F

BASE COUNT 173 a 168 c 141 g 128 t
ORIGIN

```
1  gggccctcta ggggtgcctgc tgcaggaagc acagcatagg tccagggagc ctctaattta
61 aataggagaa gtcagagctt taacagcatt gacaaaaaca agcctccaaa ttatgcaaat
121 ggaaacgaaa aagattcctc caaaggacct caatcgtctt caggtgtaaa tggtaacgtg
181 cagcctccca gtactgctgg gcagcctcct gcctctgcca tcccttctcc aagtgccagc
241 aagccctggc gcacgaagtc catgaatgtc aaacacagtg ccacctccac catgttgact
301 gtaaagcagt caagtacagc cacctcccc acaccatctt cagacagact gaaggcaacc
361 tgtctcagaa ggggtcaaaa ctgctccctc aggacagaaa tccatgcttg agaaattcaa
421 gctagtcaat gcccggactg ctttacgccc ccgcagcct cccagttcag gacctagtga
481 tggtaggaag gatgatgatg ccttttctga atctggtgaa atggaagggtt ttaacagtgg
541 tctgaatagt ggtggctcaa caaatagcag tcccaaagtg tcacctaaagt tggcccctcc
601 aaaagctgga
```


LOCUS AA495042 418 bp mRNA EST 27-JUN-1997
DEFINITION fa05f06.sl Zebrafish ICRFzfls Danio rerio cDNA clone 5D16 3'.
ACCESSION AA495042
NID g2225470
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Ostariophysi; Cypriniformes; Cyprinidae; Rasborinae;
Danio.
REFERENCE 1 (bases 1 to 418)
AUTHORS Clark,M., Lehrach,H., Johnson,S., Marra,M., Eddy,S., Hillier,L.,
Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T.,
Lacy,M., Le,N., Lennon,G., Martin,J., Moore,B., Schellenberg,K.,
Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R.
and Wilson,R.
TITLE WashU Zebrafish EST Project
JOURNAL Unpublished (1997)
COMMENT
Contact: Steve Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Steve Johnson lab internal ID - P2_60 NOTE - For this library, the
CLONE id field represents a position identifier on the original
cDNA library preparation plate. cDNA Library Preparation: Matthew
Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by:
Washington University Genome Sequencing Center Clone distribution:
Genome Systems, St. Louis, and Max Planck Institut fuer Molekulare
Genetik, Berlin Tel +49 30 84 13 1235
Seq primer: -40ml3 ET from Amersham
High quality sequence stop: 416.
FEATURES Location/Qualifiers
source 1..418
/organism="Danio rerio"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTCTTTTCTTTT3'], on
mRNA from pooled 26 somite zebrafish embryos;
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max
Planck Institut fuer Molekulare Genetik, Berlin) and was
not biochemically normalised. 70,000 clones from this
library were arrayed on high density filters and
subsequently screened by oligonucleotide hybridization
fingerprinting to identify unique or minimally redundant
clones for more intensive analysis."

205/270

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCT
TCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTCTCGACGAAGCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGA

pCB201 insert = U4

U4 ORF

R K L V E S D S D I N A N K E E L L R V L D V V P K L V Y H L H T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTCTGTCTGTGCCATTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACCT
AGGAACTCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAGACAGCACAGGTAACCGTAACCTCTGAAGGCCCTGGACCAAGTAAC TGGA

pCB201 insert = U4

U4 ORF

F L E K H S T S D F L I G P C F F L S C P I G I E D F R T V F I D L

GTGGAACAACCTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTC
CACCTTGTTGAGATAGTAAGGGATAGATGTCTTCCTCGGTTCCTACCCTATTCCAGGTACCTGTCTTTCGACGAACCCCTCTGGGTCACCTTACCCAG

pCB201 insert = U4

U4 ORF

W N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D P V E W V

CGGGACACACTTCCTGGCCATCAGCCCAACAAGACCAATCAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGECTCACCTCCG
GCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCGTAACGGAGTGGAGGGC

pCB201 insert = U4

U4 ORF

R D T L P W P S A Q Q D Q S K L Y H L P P P T V G P H S I A S P F

AGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACACATTGAGTC
TCCTATCTCTGTCAGTTTCTGTCTGTGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGGTTGATGTAAC TGA

pCB201 insert = U4

U4 ORF

E D R T V K D S T P S S L D S D P L M A M L L K L Q E A A N Y I E S

TCCAGATCGAGAAACCATCTCGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTACCCCCGGACAGCAGAAGCTGGCATCAGCTA
AGGCTATGCTCTTTGGTAGGACCTGGGTTGGAAGTCCGTTGTGAAATCCCAAGCCGTTAGTGACAGTGGGGGCTGTCTGCTTTGCGACCGTAGTCTGA

pCB201 insert = U4

U4 ORF

P D F E T I L D P N L Q A T L . G F G N H C H P R T A E R S H Q L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

TCTTAGCTCCTCCTCCTCCTCCTCTTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGT
AGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGCGGGTCTCCTCTTGCTCCTCCTCCTCTACTTTCTCCTCCTGTCCAA 2300

pCB201 insert = U4

S L L L S P L L F Q S T G S P A P G G E Q E G G G D E R G G T G

CTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTGTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACT
GAACCACGACATGGAACCTTTGAAGGATCCTTCTTACCACCCACCGCAACCTTGAACACGGGGATTGTGTAAATGACCGGAGGAGATTACTGA 2400

pCB201 insert = U4

S V C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TTGGGAAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTTCAATTACAACTCCTGGGCTTTCTGGGGAGGGGTTGAGAAAACATCAAAACACTGG
AACCCCTTTTCTACTAAGACCCAGAAAGGGAACCTGAAGAACAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCAAGTCTTTTGATGTTTGTGACG 2500

pCB201 insert = U4

V G K D D S G S F P . L L V S I T N S V A F V G G V Q K T S K H C

AGCAGTTCCTAAATGATTCTCACAAGCAACCTGAGAGAGACAGTCTTGTGAGGGAGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCAGACCC
TCGTCAGGATTTACTAAGAGTGTTCGTGGGACTCTCTCTGTGACAACACTCCCTCTAGACCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGTCTGGG 2600

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I W G R Q E A P Q I F S Q T

TTCCCAATTCATCACCCTGCCAACACTCGTCCGGAATCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG
AAGGTTAAGGTAGTGGTGACGGTTGTGAGCAGGCTTAAGACGCTATAGGTCGTGTACCGCCGGCGAGCTCAGATCTCCCGGGCAAATTTGGGCGAC 2700

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q W R P L E S R G P V . T R .

ATCAGCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCCTTCTTGACCTTGAAGGTGCCACTCCCACTGTCTTTCC
TAGTCGGAGCTGACACGGAAGATCAACGGTCCGTAGACAACAAACGGGAGGGGACACGGAAGGAAC TGGGACCTTCCACGGTGAGGGTGACAGGAAAGG 2800

S A S T V P S S C Q P S V V C P S P V P S L T L E G A T P T V L S

TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTTCTGGGGGTGSGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATTATTTTACTCCTTTAAGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCAACCCACCCGTCCTGTGCTTCCCCCTCCTAACCCCTTCTGT 2900

. . N E E I A S H C L S R C H S I L G G G V G Q D S K G E D W E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGGC
TATCGTCCGTACGACCCCTACGCCACCGAGATACCSAAGAC TCCGCC TTTCTTGGTGCACCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCCGCG 3000

N S R H A G D A V G S M A S E A E R T S V G S R G Y P H A P C S G A

ATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCTGCTTTCTTCCCTTCTTTCTCGG
TAATTCGCGCCGCCACACCAATGCGCGTCCGACTGGCGATGTGAACGCTCGCGGGATCGCGGGGAGGAAAGCGAAAGGAAGGAAAGAGAGCGG 3100

L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACSTTCGCGGGCTTTCCCGTCAAGCTCTAAATCGGGSCATCCCTTTAGGGTTCGGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGG
TGCAAGCGGCCGAAAGGGGAGTTCGAGATTTAGCCCGTAGGGAAATCCCAAGGCTAAATACAGAAATGCCGTGGAGCTGGGGTTTGTGAACTAATCC 3200

T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D .

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GTGATGGTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
320
CACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGAAATGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACC
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G
AACAACACTCAACCTATCTCGGTCTATTCTTTGATTTATAAGGGATTTTGGGGATTTTCGGCCTATTGGTTAAAAATGAGCTGATTTAACAAAAATTT
340
TTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCTAAAACCCCTAAAGCCGGATAACCAATTTTTACTCGACTAAATTGTTTTTAA
T T L N P I S V Y S F D L . G I L G I S A Y W L K N E L I . Q K F
AACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCGAGGCTCCCGAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAATG
360
TTGCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTAGGGGTCCGAGGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCA
N A N . F C G M C V S . G V E S P Q A P Q A G R S M Q S M H L N .
CAGCAACCGAGGTGTGGAAGTCCCGAGGCTCCCGAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAATGTCAGCAACCATAGTCCCGCCCTAAGTCC
380
GTCGTTGTGTCACACCTTTAGGGGTCCGAGGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGCGGGGATTGAGS
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P
GCCCATCCCGCCCTAAGTCCCGCCAGTTCGCGCCATTCGCCCCATGGCTGACTAATTTTTTTTATTTATGTCAGAGGCGGAGGCCGCTCTGCTCT
390
CGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAAAAAAATAAATACGTCCTCCGGCTCCGGCGAGACGAGAGA
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S
GAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAGCTCCCGGAGCTTGATATCCATTTTCGGATCTGATCAAGAGA
390
CTGATAAGGTTCTTCATCACTCCCGAAAAAAGCTCCGGATCCGAAAAAGCTTTTCGAGGGCCCTCGAACATATAGGTAAAGCCTAGACTAGTTCTCT
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E
CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGAGGTTCTCCGCGCGCTTGGGTGGAGAGGCTATTCCGGCTATGACTGGGCACACCA
390
GTCTACTCTAGCAAGCGTACTAAGTTTCTACCTAACGTGCGTCCAAGAGGCGGGCAACCCACCTCTCCGATAAGCCGATAGTACCCCGTGTGT
T G . G S F R M I E Q D G L H A G S P A A W V E R L F G Y D W A Q Q
GACAATCGGCTGCTCTGATCGCGCGGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTCTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAAGT
400
CTGTAGCGGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCGCGGGGCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGAC
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L
CAGGACGAGGACGCGGCATCTGTGGCTGGCCACGACGGGCTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT
410
GTCTGCTCGCTCGCGCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGATA
G D E A A R L S W L A T T G V P C A A V L D V V T E A G R D V L L
TGGGGAAGTGCCGGGCAGGATCTCTGTGATCTACCTTGCTCTGCCGAGAAATATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGA
420
ACCGGCTTCAGGCCCCGTCTTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTTCATAGGTAGTACCGACTACGTTACGCCCGGACGATGCGGAC
L G E V P G Q D L L S S H L A P A E K V S I M A D A M R R L H T L D
TCCGGCTACCTGCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGCGGCTCTTGTCGATCAGGATGATCTGGACGA
430
AGGCGGATGAGCGGTAAAGCTGGTGGTTTCGCTTTGAGCGTAGCTCGCTCGTCATGAGCCTACCTTCGGCCAGAACAGCTAGTCTACTAGACCTGCT
P A T C P F D H Q A K H R I E R A R T R M E A G L V D Q D D L D E
GAGCATCAGGGGCTCGCGCCAGCCGAACGTTTCGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCTGCTTGC
440
CTGTAGTCCCGGAGCGGTCGGCTTGACAAGCGGTCCGAGTTCCGCGGTACGCGCTCGCTCTAGAGCAGCACTGGGTACCGCTACGAGCAAGC
E H G G L A P A E L F A R L K A R M P D G E D L V V T H G D A C L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

CGAATATCATGGTGGAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTTACC GGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCTGGCGATAGTCTGTATCGCAACCSATGGGCACCT 4500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D
TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACC GCCCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 4600
I A E E L G G E W A D R F L V L Y G I A A P D S Q R I A F Y R L L
GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTTCGATTCCACCGCCGCCCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 4700
D E F F . A G L W G S K . P T K R R P T C H H E I S I P P P P S M
AAGGTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATT
TTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 4800
K G W A S E S F S G T P A G . S S S A G I S C V S S S P T P T C L L
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTCTTTTCTACTGCACTTCTAGTTGTGGTTTGCCAAACTCATCA
CGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGT 4900
Q L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S
ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTCTGTGTGAAATTGTTATCCGCTCACAAT
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACC GCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGCGGAGTGTTA 5000
M Y L I M S V Y R R P L A R A W R N H G H S C F L C E I V I R S Q
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
AGGTGTGTTGTATGCTCGCCCTTCGTATTTACATTTTCGGACCCACGGATTACTCCTCGATTGAGTGTAATTAACGCAAC 5082
F H T T Y E P E A . S V K P G V P N E . A N S H . L R V

209/270

FIG. 14.

v360 v370 v380 v390 v400 v410 v420 v430 v440
 SSVGTDTVEGFAHPAFHTLFAHNSSEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAQDLQDLRETMHMQLEVDLLKAENDRLKVAP
 SSVGT:VTE.PAH:.PHTLFAHNSSEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAQDLQDLRETMHMQLEVDLLKAENDRLKVAP
 SSVGTEVTETPAHSVPHTLFAHNSSEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAQDLQDLRETMHMQLEVDLLKAENDRLKVAP
 ^10 ^20 ^30 ^40 ^50 ^60 ^70 ^80 ^90

v450 v460 v470 v480 v490 v500 v510 v520 v530
 GPSSGSTPGQVPGSSALSSPRSLGLALTESFGPSLADTDLSPMDGISTCGFKKEVTLRVVVRMPPQHIKGDLDKQOEFLGCSKVSGKV
 GPSSG.TPGQVPGSSALSSPRSLGLAL:H:F:PSL:DTDLSPMDGISTCG:KEEVTLRVVVRMPPQHIKGDLDKQOEFLGCSKVSGKV
 GPSSGCTPGQVPGSSALSSPRSLGLALSHPPSPSLTDTDLSPMDGISTCGSKKEVTLRVVVRMPPQHIKGDLDKQOEFLGCSKVSGKV
 ^100 ^110 ^120 ^130 ^140 ^150 ^160 ^170 ^180

v540 v550 v560 v570 v580 v590 v600 v610 v620
 DWKMLDEAVPQVFKDYISKMDPASTLGLSTESIHGYSISBVKRVLDAPPEMPPCRRGVNNISVSLKGLKCKCVDSLVPETLIPKPMHQH
 DWKMLDEAVPQVFKDYISKMDPASTLGLSTESIHGYS:SBVKRVLDAPPEMPPCRRGVNNISV:LKGLKCKCVDSLVPETLIPKPMHQH
 DWKMLDEAVPQVFKDYISKMDPASTLGLSTESIHGYSLSBVKRVLDAPPEMPPCRRGVNNISVALKGLKCKCVDSLVPETLIPKPMHQH
 ^190 ^200 ^210 ^220 ^230 ^240 ^250 ^260 ^270

v630 v640 v650 v660 v670 v680 v690 v700 v710
 YISLLKRRRLVLSGPGSGTGKTYLTNRLAAYLVERS GREVTGIVSTFNMEQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
 YISLLKRRRLVLSGPGSGTGKTYLTNRLAAYLVERS GREVT:GIVSTFNMEQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
 YISLLKRRRLVLSGPGSGTGKTYLTNRLAAYLVERS GREVTGIVSTFNMEQQSCKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
 ^280 ^290 ^300 ^310 ^320 ^330 ^340 ^350 ^360

v720 v730 v740 v750 v760 v770 v780 v790 v800
 GSISELVNGALTCKYHKCPYIIGITNQPVKMTNBNGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESSDSDINAKKELLRLVLDWVPKLW
 GSISELVNGALTCKYHKCPYIIGITNQPVKMTNBNGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESSDSD:NANKKELLRLVLDWVPKLW
 GSISELVNGALTCKYHKCPYIIGITNQPVKMTNBNGLHLSFRMLTFSNNVEPANGFLVRYLRRKLVESSDSDVNANKKELLRLVLDWVPKLW
 ^370 ^380 ^390 ^400 ^410 ^420 ^430 ^440 ^450

v810 v820 v830 v840 v850 v860 v870 v880 v890
 YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTHFIDLWNNIIIPYLQEGAKDGKVBGQKAAWEDPVEWVRDTLPWPSAQDQSKLYB
 YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTHFIDLWNNIIIPYLQEGAKDGKVBGQKAAWEDPVEWVRDTLPWPSAQDQSKLYB
 YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTHFIDLWNNIIIPYLQEGAKDGKVBGQKAAWEDPVEWVRDTLPWPSAQDQSKLYB
 ^460 ^470 ^480 ^490 ^500 ^510 ^520 ^530 ^540

v900 v910 v920 v930 v940 v950 v960
 LPPFTVGFHSIASPPEDRTVKDSTPSSLDSDPLMAMLLKQEAANYIESPORETILDPNLQATL
 LPPP:VGPES.ASPPEDRTVKDSTP:SLDSDPLMAMLLKQEAANYIESPORETILDPNLQATL
 LPPFVGFHSIASPPEDRTVKDSTPSSLDSDPLMAMLLKQEAANYIESPORETILDPNLQATL
 ^550 ^560 ^570 ^580 ^590 ^600

210/270

FIG. 15.

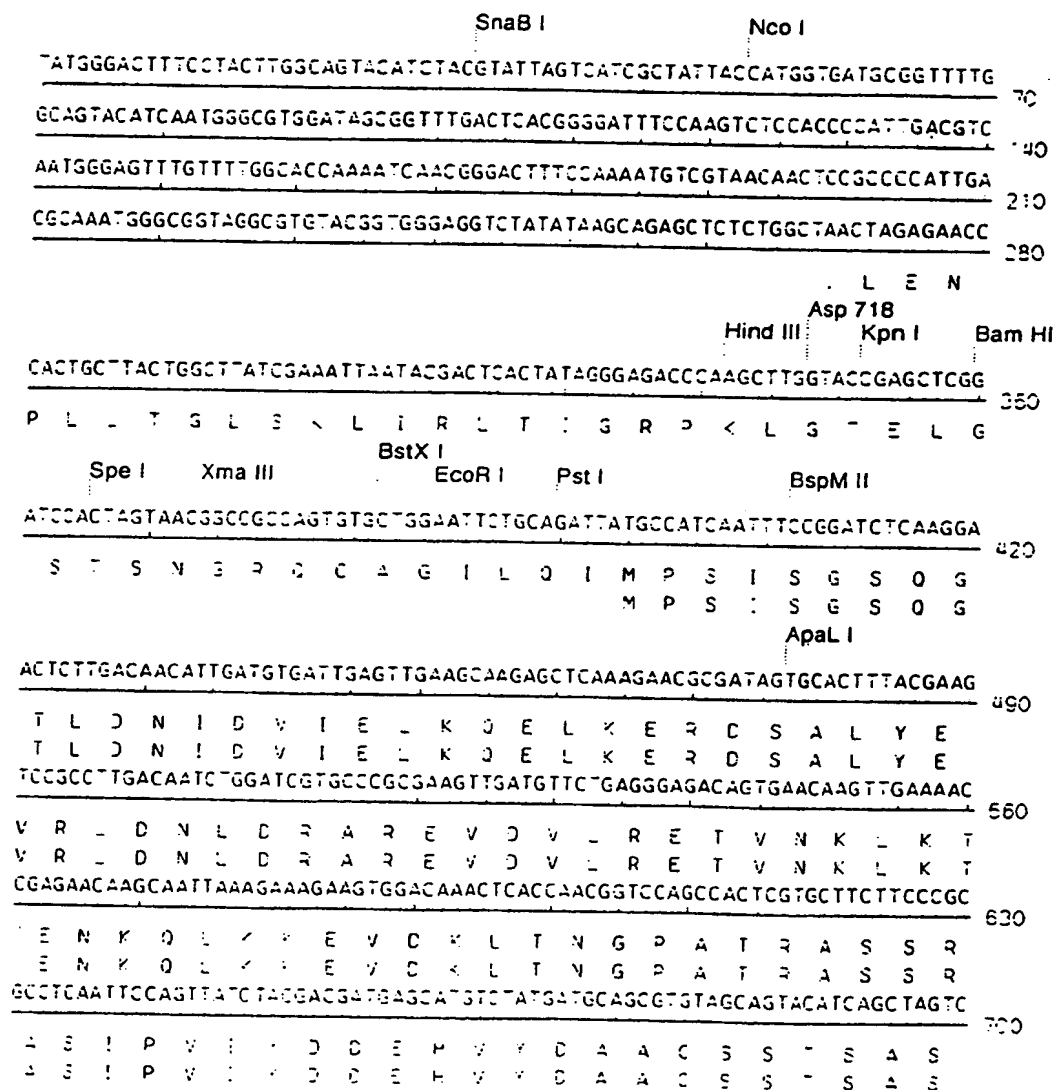


FIG. 15 CONTINUED.

Asu II

AATCTTCGAAACGATCCTCTGGCTSCAACTCAATCAAGGTTACTGTAAACGTGCACATCGCTGGAGAAAT 770
C S S K R S S S G C N S I K V T V N V D I A G E :
C S S K R S S S G C N S I K V T V N V D I A G E :

Pvu I

Hpa I

EcoR V

CAGTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCATGC 840
S S I V N P D K E I I V G Y L A M S T S Q S C
S S I V N P D K E I I V G Y L A M S T S Q S C
TGGAAAGACATTCATGTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAAATGATGTGGAGCATC 910
W K D : D V S I L G L F E V Y L S R I C V E H
W K D : D V S I L G L F E V Y L S R I C V E H

Cla I

Mlu I

AACTTGGAAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACCTCGACCCCTCATTCGAGA 980
C L G : D A R D S I L G Y Q I G E L R R V I G D
C L G : D A R D S I L G Y Q I G E L R R V I G D
CTCCACAACCATGATAACCAGCCATCCAACCTGACATTCTTACTTCCTCAACTACAATCCGAATGTTCTATG 1050
S T T M : T S H P T D : L T S S T T I R M F M
S T T M : T S H P T D : L T S S T T I R M F M
CACGCTGCCGACAGAGTCGCGTAGACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAAATGATTCCTC 1120
H G A A Q S R V D S L V L D M L L P K Q M I L
H G A A Q S R V D S L V L D M L L P K Q M I L
AACTCGTCAAGTCAATTTGACAGAGAGACGCTCTGGTGTAGCTGGAGCAACTGGAATGGAAAGAGCAA 1190
Q L V K S : L T E R R L V L A G A T G I G K S K
Q L V K S : L T E R R L V L A G A T G I G K S K

Asu II

ACTGGCGAAGACCCCTGGCTGCTTATGTATCTATTGCAACAAATCAATCCGAAGATAGTATTGTTAATATC 1260
L A K T L A A V V S I P T N Q S E D S : V N I
L A K T L A A V V S I R T N Q S E D S : V N I
Bsm I Bgl II

AGCAATTCCTGAAAACAATAGAGAAGAAATTCCTTCAAGTGGAAACGACGCTGGAAAAGATCTTGAGAAGCA 1330
S I P E N N K E E L L O V E R R L E : L R S
S I P E N N K E E L L O V E R R L E : L R S

212/270

FIG. 15 CONTINUED.

Ava III
Nsi I Xba I

AAGAATCATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTGTGTGATCCGTTTTCG 1400
 K E S C I V : L D N I P K N R : A F V V S V F A
 K E S C I V : L D N I P K N R I A F V V S V F A
 EcoR V

AAATGTCCTCACTTCAAAACAACGAAGGTCCATTGTAGTATGCACAGTCAACCGATATCAAAATCCTGAG 1470
 V V P L C N N E G P F V V C T V N R Y C I P E
 V V P L C N N E G P F V V C T V N R Y C I P E
 CTTCAAAATTCACCACAATTTCAAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCTACGTTACC

L Q I H H N F K M S V M S N R L E G F : L R Y 1540
 L Q I H H N F K M S V M S N R L E G F : L R Y
 TCCAGCAGCGGGCGGTAGAGGATGAGTATCGTCTAAGTGTACAGATGCCATCAGAGCTCTCAAAATCAT
 L R R R A V E D E Y R L T V D M P S E L F K I : 1610
 L R R R A V E D E Y R L T V D M P S E L F K I :

EcoR I

TGACTCTTCCCAATAGCTCTTCAGGCCGTCAATAATTTTATTGAGAAAACGAATTCTGTGATGTGACA 1680
 D F F P : A L Q A V N N F I E K T N S V D V T
 D F F P : A L Q A V N N F I E K T N S V D V T

Bam HI

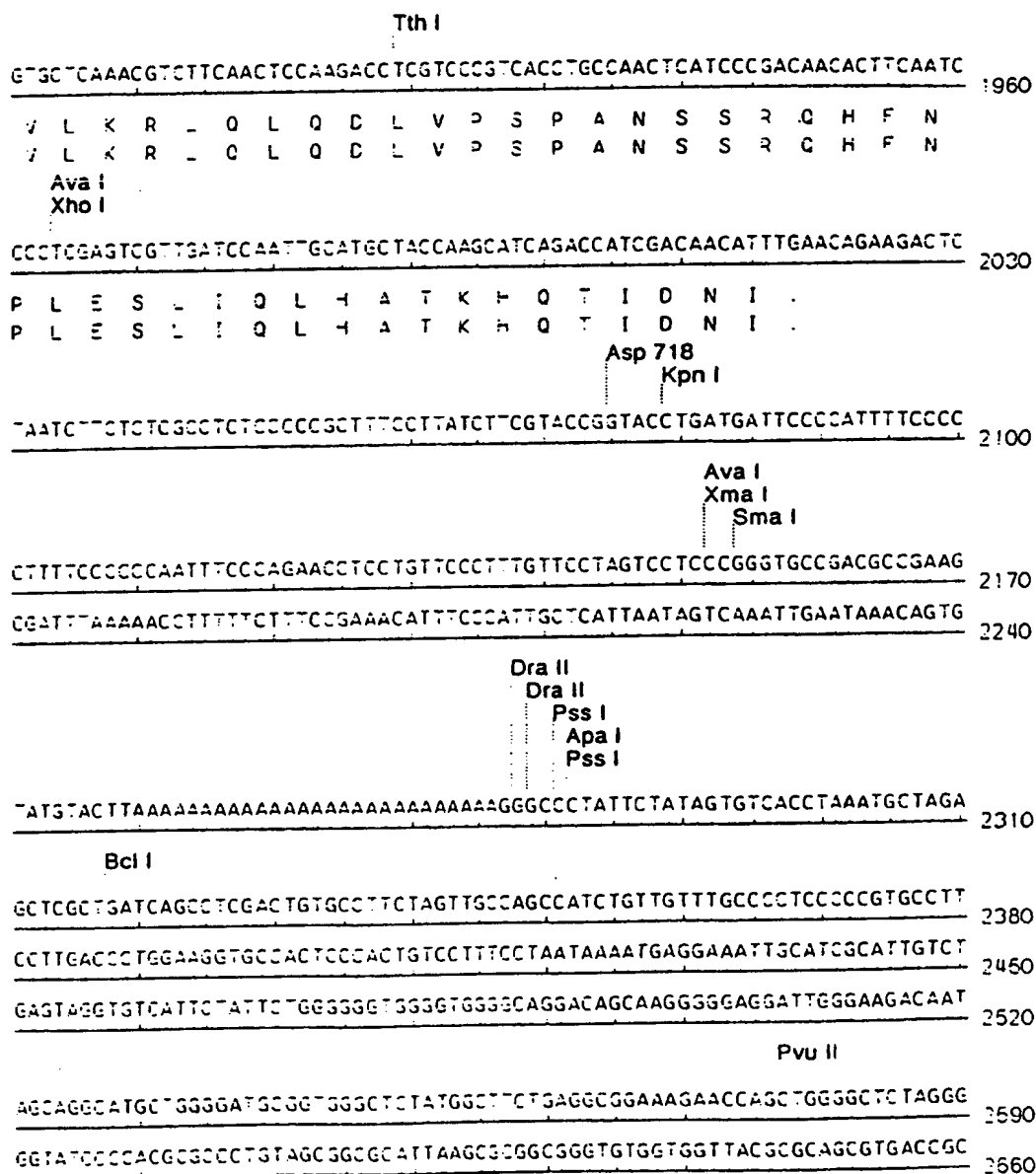
GTTGGTCCAAGAGCATGCTTGAAGTGTCTCTAACTGTCTGATGGATCCCGTGAATGGTTCATTGATTGT 1750
 V G P R A C L N C P L T V D G S R E W F : R L
 V G P R A C L N C P L T V D G S R E W F : R L
 GGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAACCTTCGGTTCGCTGCAC

W N E N F : P Y L E R V A R D G K K T F G R C T 1820
 W N E N F : P Y L E R V A R D G K K T F G R C T
 Bam HI Tth I

TTCCTTCGAGGATCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGA AAAACCCGAGAAT 1890
 S F E D P T C : V S K K W P W F D S E N D E N
 S F E D P T C : V S K K W P W F D S E N D E N

213/270

FIG. 15 CONTINUED.



214/270

FIG. 15 CONTINUED.

Nae I
TACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTGCGTTTCTCCCTTCCTTTCTCGCCACGTTGCGCCGGC 2730
TTCCCCGTCGAAGCTCTAAATCGGGGCATCCCTTAGGGTTCCGATTAGTGCTTTACGGCACCTCGACC 2800

Dra III
CCAAAAAATTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTT 2870
GACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCGAACTGGAACAACACTCAACCTATCTCG 2940
GTCTATTCTTTGATTATAAGGGATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAAC 3010
AAAAATTAAACGCGAATTAACTCTGTGGAATGTGTGTCAGTAGGGTGTTGGAAAGTCCCCAGGCTCCCCA 3080

Ava III
Nsi I
GGCAGGCAGAAGTATGCAAAGCATGCACTCTCAATTAGTCAACAACCAAGGTGTGGAAAGTCCCCAGGCTCC 3150

Ava III
Nsi I
CCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAACAACCAAGTATCCGCCCCCTAACCTCCGC 3220

Nco I
CCATCCCGCCCTTAAGTCCGCCCCAGTTCGCCCCATTCGCCCCCATGGCTGACTAATTTTTTTTATTTA 3290

Stu I
Avr II
TGCAGAGGCCGAGGCCGCTCTGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCITTTTGGAGGCC TA 3360

Ava I
Xma I
Sma I
Bcl I
GGCTTTTSCAAAAAGCTCCCGGGAGCTTGATATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGAT 3430

Xma III
GGTTTCCCATGATTGAACAAGATGGATTGCACGCGGGTCTCCGGCCGCTTGGGTGGAGAGGCTATTCCGG 3500

Nar I
Bbe I
CTATGACTGGGCAACACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCGGCTGTACGCGCAGGGGGCGC 3570

215/270

FIG. 15 CONTINUED.

Pst I

CCGGTCTTTTSTCAAGACCGACCTGTCCGGTCCCTGAATGAAGTGCAGGACGAGGCAGCGCGCTAT 3640

Bal I Fsp I Pvu II Tth I

GCTGGCTGGCCACGACGGGCGCTTCTTGGCGAGCTGTGCTCGACGTTGTCACTGAAGCGGSAAGGGACTG 3710

GCTGCTATTGGGCGAAGTGGCGGSSCAGGATCTCTGTGATCTCACCTTCTCTCTGGGAGAAAGTATCC 3780

ATCATGGCTGATGCAATGCGGGCGGTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGA 3850

AACATCGCATCGAGCGAGCACCTACTCGGATGGAAGCCGCTCTTGTGATCAGGATGATCTGGACGAAGA 3920

BssH II

GCATCAGGGGCTCGCGCCAGCGGAACTGTTCCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTC 3990

Nco I

GCTCTGACCCATGGCGATGCCCTGCTTGGCGAATATCATGCTGGGAAATGGCGGCTTTCTGATTTCATCG 4060

Nae I Rsr II

ACTGTGGCCGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATTTGCTGAAGA 4130

GCTTGGCGCGAATGGGTGACCGCTTCTCTGCTGCTTACGGTATCGCCGCTCCCGATTCGACGCGCATC 4200

Asu II

GCCTTCATCGCTTCTTGACGASTTCTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGAC 4270

GCCCAACCTGCCATCAGGAGATTTCGATTCCACCGCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTT 4340

Nae I

TTCCGGGACGCGGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTTCGCCACCCCAACT 4410

TGTTTATTGAGCTTATAATGCTTACAAATAAAGCAATAGCATCACAAATTCACCAATAAAGCATTTT 4480

Bsm I Sal I

TTCACTGCATCTAGTGTGCTTGTCCCAACTCATCAATGATCTTATCATGCTTGTATACCGTCGACC 4550

TCTAGCTAGAGCTTGGGTAAATCATGCTATAGCTGCTTCTGTGTGAAATTTGTTATCGGCTCACAATTC 4620

CACACAACATACGAGCGGGAAGCATAAAGTGTAAAGCTTGGGGTGCCTAATGAGTGAAGCTAATCAGATT 4690

FIG. 15 CONTINUED.

Pvu II

AATTGCGTTGCGCTCACTTCCCTCTTCCAGTCGGGAAACCTGTCTGCGAGCTGCATTAAATGAATCGGC 4760
CAACGCGCGGGGAGAGGCCTTTCCTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCCT 4830
CGGTGCTTCGGCTGCGGCGAGCGCTATCAGCTCACTCAAAGCGCGTAATACGCTTATCCACAGAATCAGG 4900
GGATAACGCGAGGAAAGAACAATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCCGCTTG 4970
CTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGAGCTCAAGTCAGAGGTGGC 5040
GAAACCCGACAGGACTATTAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTSCGCTCTCTTGTCC 5110
GACCTTGGCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCA 5180

ApaI

CGCTGTAGGTATCTCAGTTCGGTGTAGGTCTGTTCTCTCCAAAGCTGGGTGTGTGACGAACCCCCGTTC 5250
AGCCCGACCGCTGCGCTTATCCGTAACATATCTCTTGAGTCCAACCGGTAAGACAGGACTTATCGCC 5320

AluI

ACTGGCAGCGCCACTGGTAACAGGATTAGCAGAGCGAGSTATGTAGGCGGTGTACAGAGTCTTGAAG 5390
TGGTGGCTTAACATCGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCGAGTTACCT 5460
TCGGAAAAAGAGTTGTAGCTCTTGAATCGGCAAAACAAACACCGCTGGTAGCGGTGGTTTTTTGTTTG 5530
CAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGTCTGAC 5600

BspH I

GCTCAGTGGAAACGAAAACTCAGCTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA 5670
TCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACCTTGGTCTGACAGTTA 5740
CCAATGCTTAATCAGTGAGGCACTATCTCAGCGATCTGTCTATTTCGTTCTATCTATAGTTGCTTGACTC 5810
CCGCTGCTGTAGATAACTACATACGGGAGGGGTTACCATCTGGCCCCAGTGTCTCAATGATACCGCGAG 5880
ACCCACGCTCACCGGCTCCAGATTATTCAGCAATAAACCAGCCAGCGGGAAGGGCTGAGCGCAGAAGTGG 5950
TCCTGCAACTTTATCGGCTCTATCGAGTCTATTAATTGTTGGCGGGAAAGCTAGAGTAAGTAGTTGGCA 6020

FIG. 15 CONTINUED.

Fsp I

GTTAATAGTTTGGCAACGTTTGTSCATTGCTACAGGCATCGTSGTGTCACGCTCGTGTGGTATGG 6090
CTTCATTGAGTCCGGTCCCAACGATCAAGGCAGTTACATGATCCCCATGTTGTGCAAAAAGCGT 6160

Pvu I

TAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAAGTGGCCGCGAGTGTATCACTCATGGTTATGSCA 6220

Sca I

GCACGCGATAATTCTCTTACTTTCATGCCATCCGTAAGATGCTTTCTGTGACTGGTGAGTACTCAACCA 6300
AGTCATTCTGAGAATAGTGTATGGGGCGACCGAGTTGCTCTTCCCCGGCGTCAATACGGGATAATACCGC 6370
GCCACATAGCAGAACTTTAAAGTGCATCATTTGGAACAGCTTCTTCGGGGCGAAACTCTCAAGGATC 6440

ApaL I

TTACCGCTGTGAGATCCAGTTGATGTAACCCACTCGTGCACCCAACTGATCTTCAAGCATCTTTACTT 6510
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAGAAATGCCGCAAAAAGGGAATAAGGGCGACACG 6580

Ssp I

BspH I

GAAATGTTGAACTCATACTTTCCTTTTCAATATTATTGAAGCATTTATCAGGTTATTGTCTCATG 6650
AGCGGATACATATTGAAATGTTATTAGAAAAAATAACAAATAGGGGTTCGCGGCACATTTCCCGGAAAAG 6720

Sal I

Bgl II

Sal I

TGCCACCTGACGTGACGATCGGGAGATCTCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCT 6790

AlwI

GATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTGTGTGTTGGAGGTGCTGAGTAGT3CGCGAGCA 6860
AAATTTAAGCTACACAAAGGCAAGGCTTGACCGACAATTCATGAAGAATCTGCTTAGGGTTAGGCSTT 6930

Nru I

Mlu I

Spe I

TGCGTGCTTCGGGATGTACGGCCAGATATACGGTGGACATGATTTATGACTAGTTATTAATAGTAA 7000
TCAATTACGGGCTCATTTAGTCTATAGCCCATATATGGAGTTCCGGGTACATAACTTACGGTAAATGSCC 7070
CGCGTGCTGACCGCCCAAGCAACCGCGCGGCTATGACGTCAATTAATGAGTATGTTCCCATAGTAACGCC 7140

AATAGGGACTTTCCATTTGAGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA 7210

Nde I

GTGTATCATATGCCAAGTACGGCCCTATTGACGTCAATGACGGTAAATGGCCCGGCTGGCATTATGCCC 7280
AGTACATGACC 7350

218/270

FIG. 16.

Sst I
TATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCAC 70
Ppa I Hind III Sst I Bam HI Spe I Xma III EcoR I
TATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTG 140
R P P V C V N S
Bgl II Afl III
Mlu I
CAGATCTTGGCTATCAAATTGGTGAACCTCGACGGCTCATTGGAGACTCCACAACCATGATAACCAAGCCA 210
A D L G Y Q : G E L R R V : G D S T T M : T S H
M : T S H
TCCAACGTACATTCTTACTTCTCAACTACAATCCGAATGTTTCATGCACGGTGGCCGACAGAGTCGCSTA 280
P T D : L T S S T T I R M F M H G A A Q S R V
P T D : L T S S T T I R M F M H G A A Q S R V
GACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACCTCGTCAAGTCAATTTGACAG 350
D S L V L D M L L P K Q M : L Q L V K S : L T
D S L V L D M L L P K Q M : L Q L V K S : L T
Bbv II
AGAGACGTCTGGTGTAGCTGGAGCAACTGGAATTGGAAGAGCAAACTGGCGAAGACCCTGGCTGCTTA 420
E R R L V L A G A T G I G K S K L A K T L A A Y
E R R L V L A G A T G I G K S K L A K T L A A Y
Asu II Bsm I
TGATCTATTGGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCTGAAAACAATAAAGAA 490
V S I R T N Q S E D S : V N : S I P E N N K E
V S I R T N Q S E D S : V N : S I P E N N K E
Xmn I Bgl II
Nsi I Xba I
GAATTGCTTCAAGTGAACGACGCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAATTCTAG 560
E L L O V E R R L E K I L R S K E S C : V I L
E L L O V E R R L E K I L R S K E S C I V I L
ATAATATCCCAAGAATCGAATTGCATTGTGTTGATCCGTTTTGCAAATGTCCCACTTCAAAACAACGA 630
D N I P : N R : A F V V S V F A N V P : Q N N E
D N I P : N R : A F V V S V F A N V P : Q N N E
EcoR V
AGGTCCATTGTAGTATGCACAGTCAACCGATATCAAAATCCCTGAGCTTCAAATTCACCACAATTTCAAA 700
S P F V V C : V N P Y Q I P E L Q I H H N F
S P F V V C : V N P Y Q I P E L Q I H H N F

FIG. 16 CONTINUED.

ATGTCAGTAATGTCGAATCGTCTCGAAGGATTATCCTACCTTACCTCCGACGACGGGCGGTAGAGGATG 770
M S V M S N R L E G F I L R Y L R R R A V E D
M S V M S N R L E G F I L R Y L R R R A V E D

Sst I

AGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCTCTTCA 840
E Y R L T V Q M P S E L F K I I D F F P I A L Q
E Y R L T V Q M P S E L F K I I D F F P I A L Q

EcoR I

GGCCGTCATAAATTTTATTGAGAAAACGAATTCGTGTGATGTGACAGTTGGTCCAAGAGCATGCTTGAAC 910
A V N N F I E K T N S V D V T V G P R A C L N
A V N N F I E K T N S V D V T V G P R A C L N

Bam HI

TGTCCTCTAACTGTGATGGATCCCGTGAATGGTTTCATTGATTGTGGAATGAGAACTTCATTCCATATT 980
C P L T V D G S R E W F : R L W N E N F I P Y
C P L T V D G S R E W F : R L W N E N F I P Y

Afl III

Bam HI

Tth I

TGGAACGTGTTGCTAGAGATGECAAAAAACCTTCGCTCCTGCACTTCCTTCGAGGATCCCAACCGACAT 1050
L E R V A R D S K K T F G R C T S F E D P T D I
L E R V A R D S K K T F G R C T S F E D P T D I

Bbv II

CGTCTCTAAAAAATGGCCGTGTTTCGATGGTGAAAACCCGAGAAATGTGCTCAAACGTCTTCAACTCCAA 1120
V S K K V P W F D G E N P E N V L K R L Q L Q
V S K K V P W F D G E N P E N V L K R L Q L Q

Tth I

Xho I

GACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCTCGAGTGGTTGATCCAATTGC 1190
D L V P S P A N S S R Q H F N P L E S L I Q L
D L V P S P A N S S R Q H F N P L E S L I Q L

Bbv II

ATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATCTTCTCTCCCTCTCCCCCG 1260
H A T K H O T : D N I
H A T K H O T : D N I
CTTTCCTTATCTCTCTACCGTACCTGATGATTCCCATTTTCCCTCTTCTCCCTCAATTTCCAGAAC 1330

Xma I

Sma I

CTCTGTTCCTTTTCTCTAGTCTCTCCCGGTGCCGACGCCGAAGCGATTAAAAACCTTTTCTTTCC 1400
Xmn I

GAAACATTTCCCATTTGCTCATTAAATAGTCAAATTSAAATAACAGTGTATGTACTTAAAAA 1470

FIG. 16 CONTINUED.

Sst I Bcl I

AAAAAAAAAAAAAGGGCCCTATTCTATAGTGTACCTAAATGCTASAGCTCGCTGATCAGCCTCGACTGTG 154C
 CTTCTAGTTGCCAGCCATCTGTGTTTGCCCCCCCCCGTCCCTTCTTGACCCGGGAAGGTGCCACTC 161C
 CCACTGTCTCTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTAGTAGGTGTCATTCTATTCTGGG 166C

Bbv II

GGGTGGGGTGGGSCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGCATGCTGGGGATGCGGTG 175C
 GGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCG 182C
 GCGCATTAAAGCGCGCGGTGTGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC 189C
 CGCTCCTTTGGCTTTCTTCCCTTCTTCTCGCCACGTTGCGCGGCTTTCCCCGTCAAGCTCTAAATCGG 196C
 GGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC TTGATTAGGGTGATG 203C

Dra III

GTTACAGTAGTGGSCCATGSCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAA 210C
 TAGTGGACTCTTGTCCAAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTTGATTATAAGGG 217C

Xmn I

ATTTTGGGGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCT 224C
 GTGGAATGTGTCTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCAT 231C

Ava III

Nsi I

GCATCTCAATTAGTCAGCAACCAGGTGTGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGC 236C

Ava III

Nsi I

ATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACCTCCGCCCATCCCGCCCCTAACCTCCGCCCA 245C
 GTTCCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCCTCTGC 252C

Stu I

Xma I

Sma I

CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGA 259C

Bcl I

GCTTGTATATCTATTTCGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATG 266C

Xma III

GATTGCACGCAAGTTCTCCGCCCCCTGCGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAAACAGACAAT 273C

Nar I

Bbe I

CGGCTGCTCTGATGCGGCGCTGTTCGGGCTGTGAGCGCAGGGGCGCCCGGCTTTTGTCAAGACCGAC 280C
 CTGTCCGGTCCCTGAATGAACGAGGACGAGGACGCGGCTATCGTGGCTGGCCACGACGGGCGTTC 287C

Fsp I

Tth I

CTTGGCAGCTGTGCTGACGTTGTACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG 294C
 GCAGGATCTCTGTATCTACCTTGCTCTGCGGAGAAAGTATCCATCATGGCTGATGCAATGCGGGCG 301C

FIG. 16 CONTINUED.

CTGCATACGCTTATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA 3080
CTCGGATGGAAGCCGGTCTTGTCATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGA 3150

BssH II

ACTGTTCCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGC 3220

Rsr II

TTGCCGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTTCGACTGTGGCCGGCTGGGTGTGGCGG 3290
ACCGCTATCAGGACATAGCGTTGGCTACCGGTGATATTCCTGAAGAGCTTGGCGGGCGAATGGGCTGACCG 3360
CTTCCTCGTGCCTTACGGTATCGCCGCTCCCGATTCCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAG 3430

Asu II

TTCTTCTGAGCGGACTCTTGGGTTGAAATGACCGACCAAGCGACGCCAACCTGCCATCAGGATTT 3500
CGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCCGAATCGTTTCCGGGACGCCGGCTGGATGATC 3570
CTCCAGCGCGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTT 3640

Bsm I

ACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTCACTGCATTCTAGTTGTGGTTT 3710

Sna I

GTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCCGACCTCTAGCTAGAGCTTGGCGTAATCA 3780
TGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAAATTCACACAACATACGAGCCGGAAGCA 3850
TAAAGTGTAAGGCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 3920

Sbo I

TTTCCAGTCGGGAAACCTGTCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGAGAGGCGGTTTG 3990
CGTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTCGGCTCGGTCTCGGCTGCGCGCAGCGG 4060

Alu III

TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTG 4130
AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTCATAGGCTCCGC 4200
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGAT 4270
ACCAGGCGTTTCCGCTGGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCGCGCGTTACCGGATACCT 4340
GTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAGCTGTAGGTATCTCAGTTGCGTG 4410

ApaL I

TAGGTCGTTCCGTCGAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCCTTATCCG 4480

Alu I

GTAACATCTCTCTGAGTCCAAACCGGTAAAGACAGCACTTATGCGCACTGGCAGCAGCCACTGGTAACAG 4550
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTTAACACGGCTACACT 4620
AGAAGSACASTATTGGTATCTGCGCTCTGCTGAAGCCASTTACCTTCGGAAAAAGAGTTGGTAGCTCTT 4690
GATCCGCAAAACAAACACCGCTTGTAGCGGTGGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAA 4760
AAAAGGATCTCAAGSAGATCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACACAGT 4830

BspH I

TAAAGGATTTTGTCTATGAGATTATCAAAAAGGATCTTCACTAGATCTTTTAAATTAATAAAGAGTT 4900

FIG. 16 CONTINUED.

TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCASTGAGGCACC 4970
TATCTCAGCGATCTGTCTATTTCTTCATCCATAGTTGGCTGACTCCCCGTCGTGTAGATAACTACGATA 5040
Ppa I

CGGGAGGGCTTACCATCTGSCCCAGTGCTSCAATGATACCGGAGACCCACGCTCACCGGCTCCAGATT 5110
TATCAGCAATAAACCCAGCCAGCCGAAGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCCTCCAT 5180
Fsp I

CCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT 5250
GCCATTGCTACAGGCATCGTGGTCTCAGCTCGTCTGTTGGTATGGCTTCATTGAGCTCCGGTCCCAAC 5320
Pvu I

GATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGT 5390
TGTGAGAAGTAAGTTGGCCGAGTGTATCACTCATGTTATGGCAGCACTGCATAATTCTCTTACTGTC 5460
Sca I Sbo I

ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGATEC 5530
GGCGACCGAGTTTCTCTTGGCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGT 5600
Xmn I

GCTCATCATTGGAAAACGTTCTTCCGGGCGSAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCC 5670
EcoK ApaL I

ATGTAACCCACTCGTGACCCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAA 5740
AAACAGGAAGGCAAAATGCCGCAAAAAAGGSAATAAGGGCGSACGSAATGTTGAATACATCACTCTT 5810
Ssp I BspH I

CCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATT 5880
TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGSAAAAGTGCCACCTGACGTGACGGATCGG 5950
Bgl II AlwN I

GAGATCTCCCGATCCCTATGGTGGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT 6020
CTGCTCCCTGCTTGTGTTGGAAGTCCGCTGAGTAGTGCAGGCAAAATTTAAGCTACAACAAGGCAAG 6090
Nru I

GCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGCC 6160
Afl III Mlu I Spe I

CAGATATACGCCTTGACATTGATTATGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTAT 6230
AGCCCATATATGGAGTTCGCGCTTACATTAACCTACGGTAAATGGCCGCGCTGGCTGACCGCCCAACGACC 6300
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA 6370
Nde I

ATGGGTGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC 6440
CCTATTGACGTCAATGACGGTAAATGGCCGCGCTGSCATTATGCCAGTACATGACCTTATGGGACTTTC 6510
SnaB I

CTACTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA 6580
TGGCGTGGATAGCGGTTTGAATCACGGGSAATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTG 6650
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTGAACAACCTCCGCCCATTTGACGCAAAATGGGCG 6720
GTAGGCGTGTACGGTGGGAGGTCTA 6745

FIG. 17.

GGTCCGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGC 70
Fsp I
CAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACAGCTCGTCGTTTGGTAT 140
GGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCG 210
Pvu I
GTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAAGTGTATCACTCATGGTTATGG 280
Sca I
CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTCTGTGACTGGTGAGTACTCAAC 350
CAAGTCATTCTGAGAATAGTCTATGCGCGGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC 420
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCAAGGA 490
ApaL I
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCAACTGATCTTCAGCATCTTTTAC 560
TTTACCGAGCGTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACA 630
Ssp I BspH I
CGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCA 700
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCGAAA 770
Sal I Bgl II Sal I
AGTSCACCTGACGTGACGGATCGGGAGATCTCCGATCCCTATGGTCGACTCTCAGTACAATCTGCT 840
AlwI
CTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAG 910
CAAAAATTAGCTACAACAAGGCAAGGCTTGACGACAAATTCATGAAGAATCTGCTTAGGGTTAGGCGT 980
Nru I Mlu I Spe I
TTTGCCTGCTTCGCGATGTACGCGCCAGATATACCGTTGACATTGATTATTGACTAGTTATTAATAGT 1050
AATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTACGGTAAATGG 1120

FIG. 17 CONTINUED.

CCCGCCTGGCTGACCGCCCAACGACCCCGCCCAATTGACGTCAATAATGACGTATGTTCCCATAGTAACG 1190
CCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATC 1260
Nde I
AAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGCTAAATGGCCCGCCTGGCATTATGC 1330
SnaB I Nco I
CCASTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG 1400
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC 1470
ACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTAACAA 1540
CTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGCTATATAAGCAGAGCTCTCTGG 1610
Hind III
CTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTT 1680
L E N D L L T G L S K L I R L T : G R P < L
Asp 718 Bam HI BstX I
Kpn I Spe I Xma III EcoR I Pst I
GGTACCGAGCTCGGATCCACTAGTAACGCGCGCCAGTGTGCTGGAATCTGCAGATCGCCGCTGCCACCT 1750
G T E L G S T S N G R Q C A G : L O I A A A T
Ava I
Xma I
Sma I
Pvu II
CAACCTTCGGACAACATTGCTAAGATCCCCGGGATACTCATCCTATTCTCCACACTTATCAGTCTCAGC 1820
S T F G Q H S L R S P G Y S S Y S P H L S Y S A
Spe I
Sal I
TGATAAGGACACAATGTCTATGCACTCACAGACTAGTCCGACGACCTTCTTCACAAAAACCAAGCTATTCA 1890
D Y D T M S M H S Q T S R R P S S Q K P S Y S
W E M H S Q T S R R P S S Q K P S Y S
GGCCAAATTCATTACCTTATCGTAAATGCCACCTTCAAGAGTTTACATCCACCGAGCACAGAAATGGCGG 1960
G Q F H S L D R K C H L Q E F T S T E F R Y A
G Q F H S L D R K C H L Q E F T S T E F R Y A

FIG. 17 CONTINUED.

Ava I Bam HI
CTCTCTT¹GAGCCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATTCTTCAGGATCCTACTCGGGCGG 2030
A L L S P R R V P N S M S K Y D S S G S Y S A R
A L L S P R R V P N S M S K Y D S S G S Y S A R
Ava I
TCCCGAGGTGGAAGCTCTACTGSTATCTATGGAGAGACGTCCAAC²TGCACAGACTATCCGATGAAAAA 2100
S R G G S S T G I Y G E T F Q L H R L S D E K
S R G G S S T G I Y G E T F Q L H R L S D E K
Bam HI Nde I
TCCCCGACATCTGTCCAAAAGT³GAGATGGGATCCCAACTATCACTGGCTAGCAGCAGCATATGGAT 2170
S P A H S A K S E M G S Q L S L A S T T A Y G
S P A H S A K S E M G S Q L S L A S T T A Y G
Sal I
CTCTCAATGAGAAGTACGAACATGCTATTTCGGGACATGGCACGTGACTTGGAGTGTTACAAGAACA⁴CTGT 2240
S L N E K Y E H A I R D M A R D L E C Y K N T V
S L N E K Y E H A I R D M A R D L E C Y K N T V
Hind III
CGACTCACTAACCAAGAAACAGGAGAAC⁵TATGGASCATTGTTTGATCTTTT⁶GAGCAAAAGCTTAGAAAA 2310
D S L T K K Q E N Y G A L F D L F E Q K L R K
D S L T K K Q E N Y G A L F D L F E Q K L R K
Cla I
CTCACTCAACACATTGATCGATCCAAC⁷TGAAGCCTGAAGAGGCAATACGATT⁸CAGGCAGGACATTGCTC 2380
L T Q H I D R S N L K P E E A I R F R Q D I A
L T Q H I D R S N L K P E E A I R F R Q D I A
ATTGAGGGATATTAGAATCATCTTGATCCAAC⁹TCAGCTCATGCTAACGAAGGCGCTGGTGAGCTTCT 2450
F L R D I S N H L A S N S A H A N E G A G E L L
F L R D I S N H L A S N S A H A N E G A G E L L
Cla I Cla I
TCGTCAACCATCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTGGTGA¹⁰AAAGCAGCAAG 2520
R G P S L E S V A S H R S S M S S S S K S S K
R G P S L E S V A S H R S S M S S S S K S S K

FIG. 17 CONTINUED.

Bam HI

CAGGAGAAGATCAGCTTGAGCTGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGT 2590

D E K I S L S S F G K N K K S W I R S S L S K
D E K I S L S S F G K N K K S W I R S S L S K

Nde I BspM II

TCACCAAGAAGAACAAGAAGTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAAGTCT 2660

F T K K K N K Y Y D E A H M P S I S G S G G T L
F T K K K N K Y Y D E A H M P S I S G S G G T L

ApaI

TGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTACGAAGTCCGC 2730

D N : D V I E L K Q E L K E R D S A L Y E V R
D N : D V I E L K Q E L K E R D S A L Y E V R

CTTGACAATCTGGATCGTGCCCGGAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGA 2800

L D N L D R A R E V D V L R E T V N K L K T E
L D N L D R A R E V D V L R E T V N K L K T E

ACAAGCAATTAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCGCGCTC 2870

N K D L K K E V D K L T N G P A T R A S S R A S
N K D L K K E V D K L T N G P A T R A S S R A S

AATTCCAGTTATCTACGACGATGAGCATGTCATGATGCCAGGTGTAGCACTACATCAGCTAGTCAATCT 2940

I P V I Y D D E H V Y D A A C S S T S A S G S
I P V I Y D D E H V Y D A A C S S T S A S G S

Asu II

TCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTT 3010

S K R S S G C N S I K V T V N V D I A G E I S
S K R S S G C N S I K V T V N V D I A G E I S

Pvu I Hpa I EcoR V

CGATCGTTAACCCTGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCACTGCTGAA 3080

S I V N P D K E I I V G Y L A M S T S D S C W K
S I V N P D K E I I V G Y L A M S T S D S C W K

AGACATTGATGTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAAITGATGTGGAGCATCAACT 3150

D : D V S I L E L F E V Y L S R I D V E H C L
D : D V S I L E L F E V Y L S R I D V E H C L

Cla I
 GGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAAGTTCGACGCGTCATTGGAGACTCCA
 3220
 G I D A R D S I L G Y Q I G E L R R V I G D S
 G I D A R D S I L G Y Q I G E L R R V I G D S
 CAACCATGATAACCCAGCCATCCAAGTACATTTCTTACTTCTCAACTACAATCCGAATGTTTCATGCACGG
 3290
 T T M I T S H P T D I L T S S T T I R M F M H G
 T T M I T S H P T D I L T S S T T I R M F M H G
 TGCCGCACAGAGTCGCGTAGACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAAGTCT
 3360
 A A Q S R V C S L V L D M L L P K Q M I L Q L
 A A Q S R V C S L V L D M L L P K Q M I L Q L
 GTCAAGTCAATTTTGACAGAGAGACGCTCTGGTGTTAGCTGGAGCAACTGGAATTGSAAGAGCAAAGTGG
 3430
 V K S I L T E R R L V L A G A T G I G K S K L
 V K S I L T E R R L V L A G A T G I G K S K L
 Asu II
 Bsm I
 CGAAGACCTTGCTGCTTATGTATCTATTCGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCAT
 3500
 A K T L A A Y V S I R T N Q S E D S I V N I S I
 A K T L A A Y V S I R T N Q S E D S I V N I S I
 Bgl II
 TCTGAAACAAATAAGAAGAAATTGCTTCAAGTGAACGACGCTTGGAAAAGATCTTGAGAAGCAAAGAA
 3570
 P E N N K E E L L O V E R R L E K I L R S K E
 P E N N K E E L L O V E R R L E K I L R S K E
 Ava III
 Nsi I
 Xba I
 TCATGCATCGTAATTTCTAGATAATATCCCAAAGAATCGAATTGCATTTTGTATCCGTTTTTGCAAATG
 3640
 S C I V I L C N I P K N R I A F V V S V F A N
 S C I V I L C N I P K N R I A F V V S V F A N
 EcoR V
 TCCCACTTCAAAACAACGAAGGTCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCA
 3710
 V F L O V N E S P F V V C T V N R Y C I P E L Q
 V F L O V N E S P F V V C T V N R Y C I P E L Q
 AATTCACCACAATTTCAAAAATGAGTAATGTGGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGA
 3780
 I H H N F K M S V M S N R L E G F I L R Y L R
 I H H N F K M S V M S N R L E G F I L R Y L R

FIG. 17 CONTINUED

CGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACT 3850
R R A V E D E Y R L T V Q M P S E L F K I I D
R R A V E D E Y R L T V Q M P S E L F K I I D
EcoR I

CTTCCCAATAGCTCTTCAGGCCGTCAATAATTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGG 3920
F F P I A L C A V N N F I E K T V S V D V T V G
F F P I A L C A V N N F I E K T V S V D V T V G
Bam HI

TCCAAGAGCATGCTTGAAGTGTCTCTAACTGTCTGATGCCGTGAATGGTTCATTGATTGTGGAAT 3990
P R A C L N C P L T V D G S R E W F I R L W N
P R A C L N C P L T V D G S R E W F I R L W N
GAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAACCTTCGGTCCGTGCACTTCTC 4060
E N F I P Y L E R V A R D G K K T F G R C T S
E N F I P Y L E R V A R D G K K T F G R C T S
Bam HI Tth I

TGAGGATCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGGAGAATGTGCT 4130
F E D P T D : V S K K W P W F D G E N P E N V L
F E D P T D : V S K K W P W F D G E N P E N V L
Tth I Ava I Xho I

CAAACGTCTTCAACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCGACAACACTTCAATCCCGTC 4200
K R L Q L Q D L V P S P A N S S R Q H F N P L
K R L Q L Q D L V P S P A N S S R Q H F N P L
GAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATC 4270
E S L I Q L H A T K H O T I D N I
E S L I Q L H A T K H O T I D N I
Asp 718 Kpn I

TTCTCTCGCCTCTCCCGCGCTTTCTTATCTCTGTAACGGTACCTGATGATTCCCGATTTTCCCCCTTT 4340
Ava I Xma I Sma I

CCCCCAATTCCGAGAACCTCTGTTCCTTTGTTCTAGTCTCCCGGGTGCCGACGCCGAAGCGATT 4410

FIG. 17 CONTINUED.

TAAAAACCTTTTCTTTCCGAAACATTTCCCATTCGTCATTAATAGTCAAATTGAATAAACAGTGATGT 4480

Dra II
Dra II
Pss I
Apa I
Pss I

ACTTAAAAAAAAAAAAAAAAAAAAAAAAAGGGCCCTATTCTATAGTGTACCTAAATGCTAGAGCTCG 4550

Bcl I

CTGATCAGCCTCGACTGTGCTTCTAGTTSCCAGCCATCTSTTGTGTTGCCCTCCCCCGTGCCTTCCTTG 4620

ACCTTGGAAAGGTGCCACTCCCACTGTCTTTCTTAATAAATGAGGAAATTGCATCGCATTGTCTGAGTA 4690

GGTGTCTTCTATTCTGGGGGTGGGTGGGCGAGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG 4760

Pvu II

GCATGCTGGGATGCGGTGGCTCTATGGCTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTAT 4830

CCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACAC 4900

Nae I

TGCCAGCGCCCTAGCGCGCGCTCTTTGCTTTCTTCCTTCTTTCTCGCCACGTTGCGCGGCTTTTC 4970

CCGTCAAGCTCTAAATCGGGCATCCCTTTAGGGTTCCGATTTAGTGTCTTACGGCACCTCGACCCCAAA 5040

Dra III

AAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT 5110

TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTCCAAACTGGAACAACACTCAACCTATCTCGGTCTA 5180

TTCTTTTGATTTATAAGGGATTTTGGGGATTTGGGCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA 5250

TTAACGCGAATTAAATCTGTGAATGTGTGTCAGTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAG 5320

Ava III
Nsi I

GCAGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGSAAGTCCCCAGGCTCCCCAGC 5390

Ava III
Nsi I

AGGCAGAAATATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCCGCCCTAACTCGCCCATC 5460

FIG. 17 CONTINUED.

CCGCCCTAACTCCGCCAGTCCGCCATTCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAG 5530
Nco I
AGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTGTGAGGAGGCTTTTGGAGGCCCTAGGCTT 5600
Stu I
Avr II
TTGCAAAAAGCTCCCGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTT 5670
Ava I
Xma I
Sma I
Bcl I
CGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGCCCGCTTGGTGGAGAGGCTATTCGGCTATG 5740
Xma III
ACTGGGCACAACAGACAACTCGGCTGCTCTGATGCCGCCGTGTTCGGCTGTACGCGCAGGGGCGCCCGGT 5810
Nar I
Bbe I
TCTTTTGTCAAGACCGACCTGTCCGGTGGCCCTGAATGAAGTGCAGGACGAGGCAGCGCGGTATCGTGG 5880
Pst I
CTGGCCACGACGGCGTTCCCTGGCGAGCTGTGCTCGACGTGTCACTGAAGCGGGAAGGGACTGGCTGC 5950
Bal I
Fsp I
Pvu II
Tth I
TATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTATCTCACCTGTCTCTGCCGAGAAAGTATCCATCAT 6020
GGCTGATGCAATGCGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACAT 6090
CGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATC 6160
BssH II
AGGGGCTCCGCCAGCGGAAGTGTTCGCCAGGCTCAAGGCGCGCATGCCCCAGCGGAGGATCTCGTCGT 6230
Nco I
GACCCATGGCGATGCTGCTTGCCTAATATCATGTTGGAATAATGGCGCTTTTCTGGATTCATCGACTGT 6300

FIG. 17 CONTINUED.

Nae I **Rsr II**

GGCCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTG 6370
GCGGCGAATGGGCTGACC GC T TCCTCGT GCTTTACGGTATCGCCGCTCCCATTTCAGCGCATCGCCTT 6440

Asu II

CTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAASCGACGCCCA 6510
ACCTGCCATCACGAGATTTCGATTCCACCGCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCG 6580

Nae I

GGACCGCGGCTGGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCGCCACCCCCAACTTGTTT 6650
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTACAC 6720

Bsm I **Sal I**

TGCATTCTAGTTGTGGTTTGCCAACTCATCAATGTATCTTATCATGTCTGTATACCGTGCACCTCTAG 6790
CTAGAGCTTGGCGTAATCATGGTCATAGCTSTTTCCGTGTGAAATTGTTATCCGCTCACAAATCCACAC 6860
AACATACGAGCCGSAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTSGAGCTAACTCACATTAATTG 6930

Pvu II

CGTTGCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAAATGAATCGGCCAACG 7000
CGCGGGGAGAGGGCGGTTCGCTATTGGGCGCTTTCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTC 7070
GTTTCGGCTGCGGCSAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA 7140
ACGCAGGAAAGAACAATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC 7210
GTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC 7280
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCC 7350
TGGCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCCTTTCTCAATGCTCACGCTG 7420

Apal I

TAGGTATCTCAGTTGGGTGTAGGTGGTTCCGTCGAAGCTGGGCTGTGTGACGGAACCCCCGTTACGCC 7490
GACCGCTCGCGCTTATCGGTAAGTATCGTCTTGAATCCAACCGGTAAGACACGAGTTATCGCCACTGG 7560

FIG. 17 CONTINUED.

AlwI

CAGCAGCCACTGCTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG 7630
GCCTAAC TACGGCTACACTAGAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA 7700
AAAAGAGTTGGTAGCTCTTGATCCGGCAAAACCAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGC 7770
AGCAGATTACCCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCA 7840

BspH I

GTGGAACGAAAACTCAGCTTAAGGGATTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT 7910
TTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT 7980
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTGCCTGACTCCCGT 8050
CGTG TAGATAAC TACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA 8120
CGCTCACC GGCTCCAGATTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT 8183

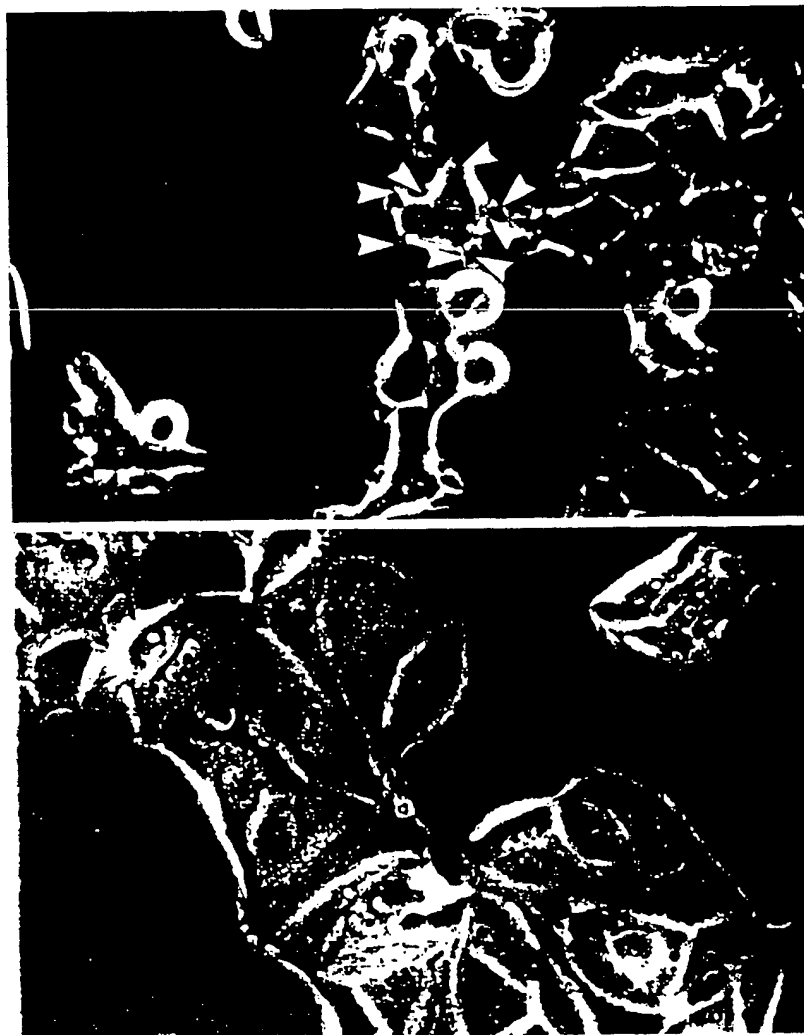


Figure 18 : Phase contrast images of MCF-7 cells transfected with pCB201 (upper) compared to mock (control) transfected MCF-7 cells (bottom).

The control cells are spread out on the tissue culture plastic and exhibit few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths (arrowhead) per cell.

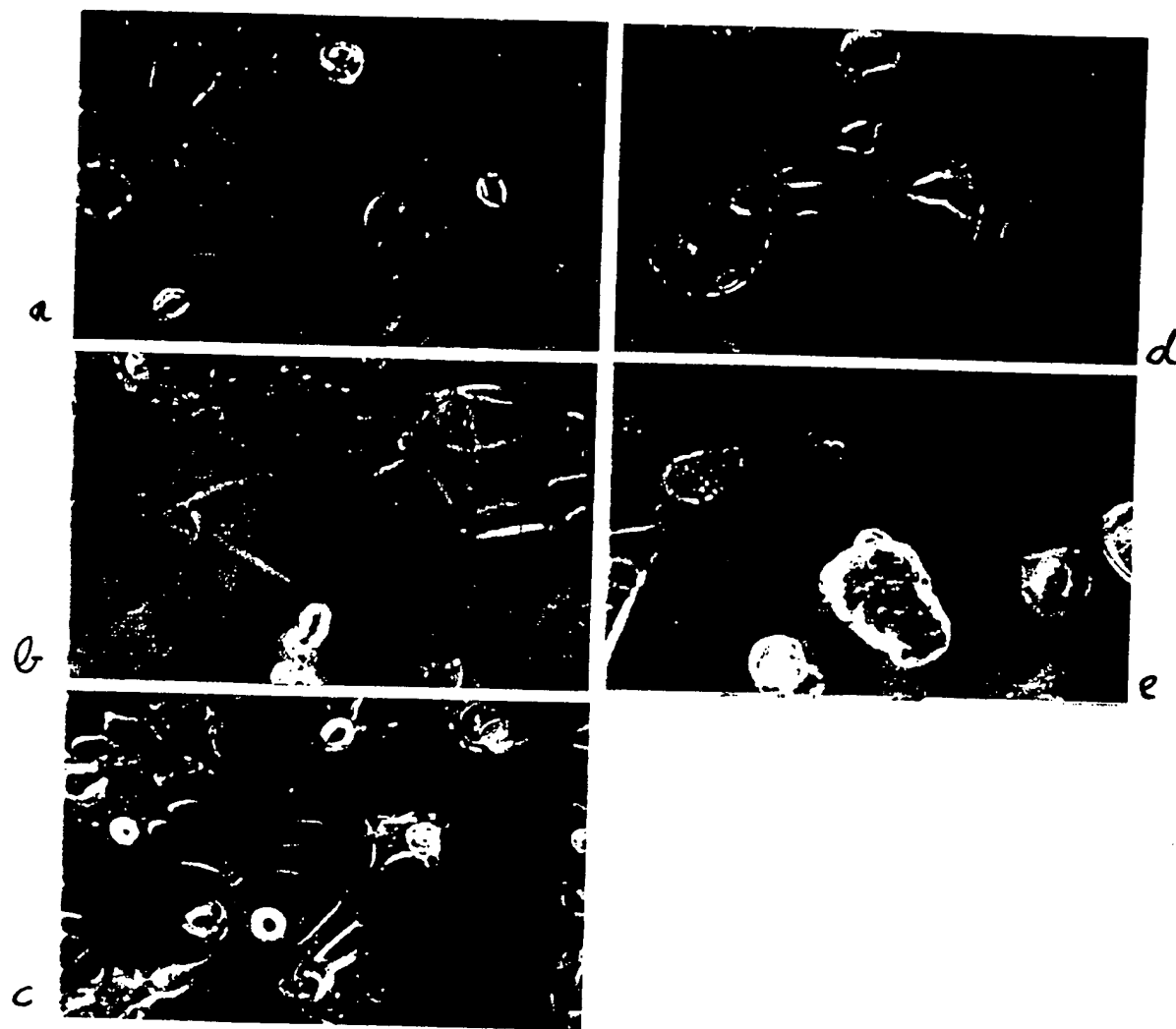


Figure 19 : Phase contrast image of MCF-7 cells, transfected with pcDNA3 (19a), pCDU4 (19b), pCDU3 (19c) pCDU2 (19d) and pTB72 (19e).

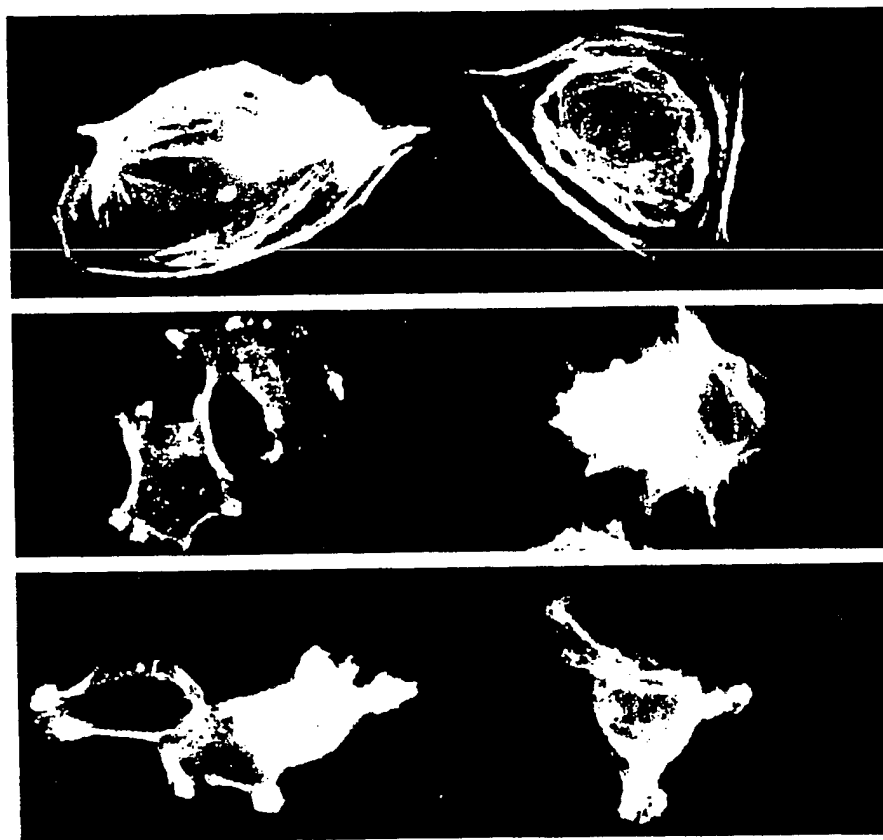


Figure 20 : F- actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

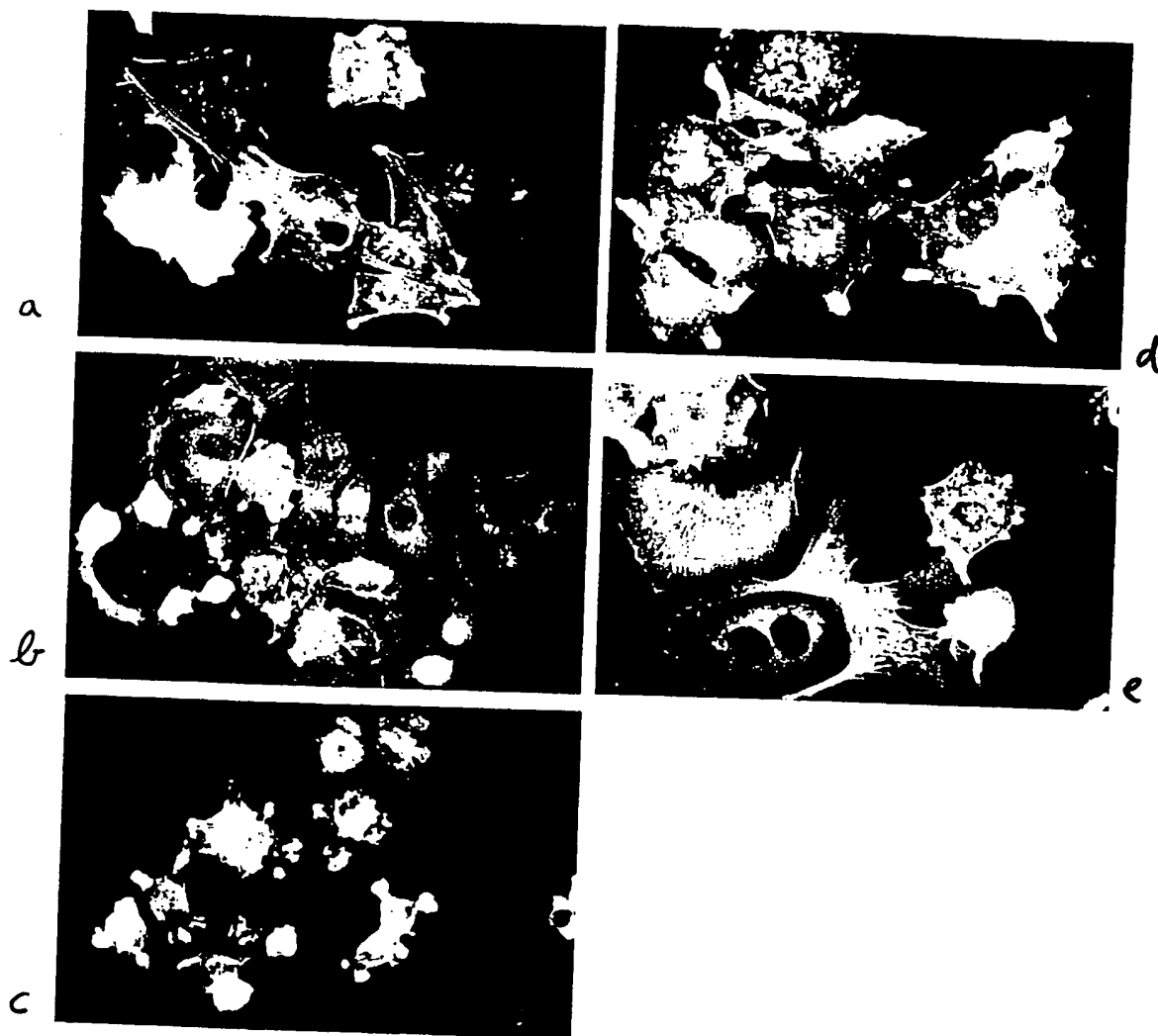


Figure 21 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3 (21a), pCDU4 (21b), pCDU3 (21c) pCDU2 (21d) and pTB72 (21e)

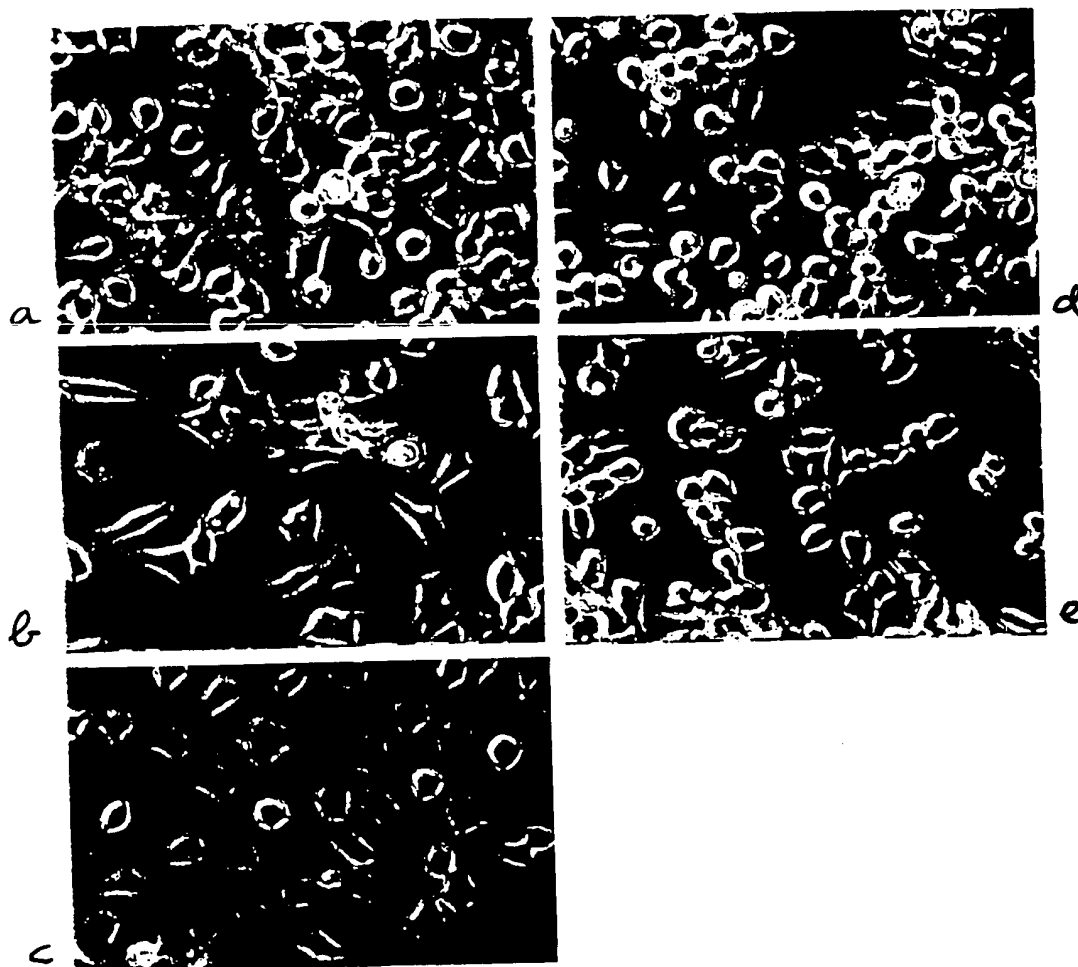


Figure 22 : Phase contrast image of N4 neuroblastoma cells, transfected with pcDNA3 (22a), pCDU4 (22b), pCDU3 (22c) pCDU2 (22d) and pTB72 (22e)

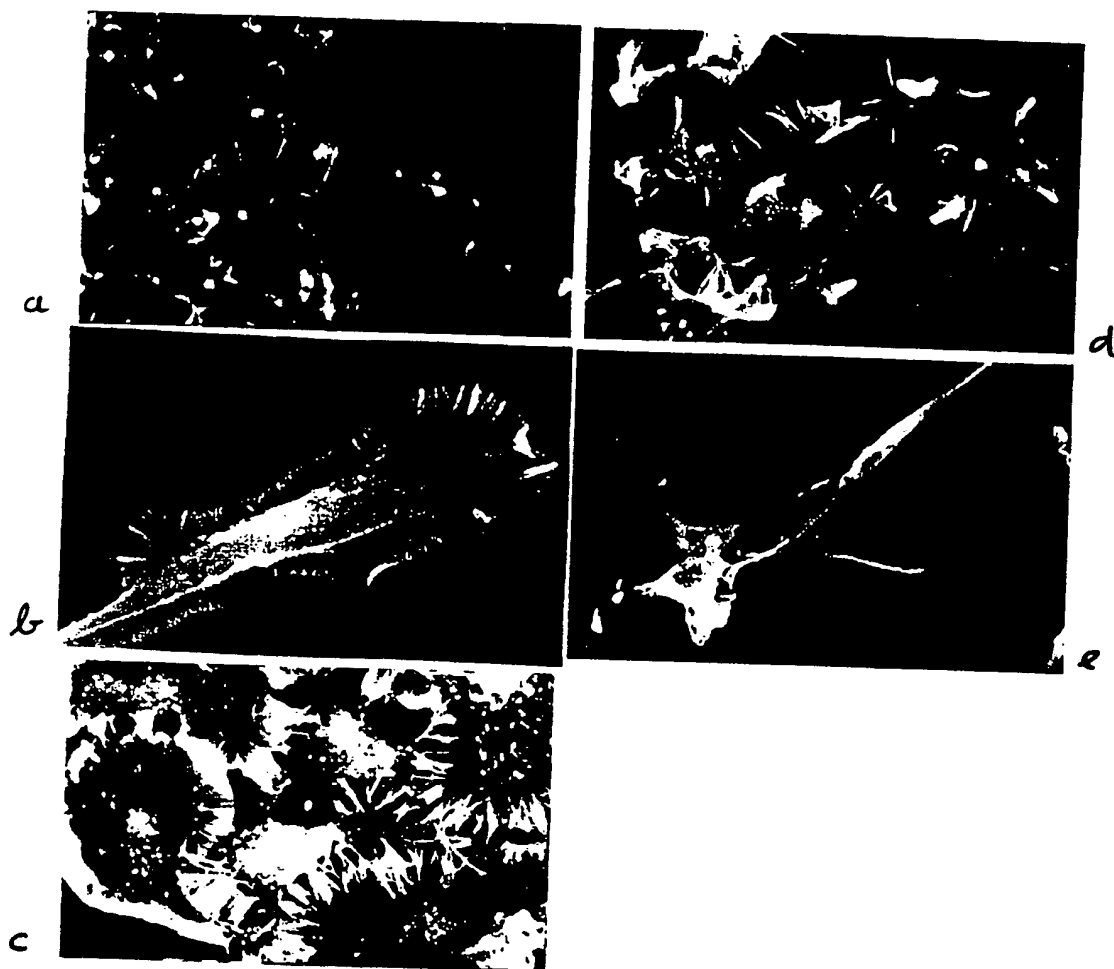


Figure 23 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pcDNA3 (23a), pCDU4 (23b), pCDU3 (23c) pCDU2 (23d) and pTB72 (23e)

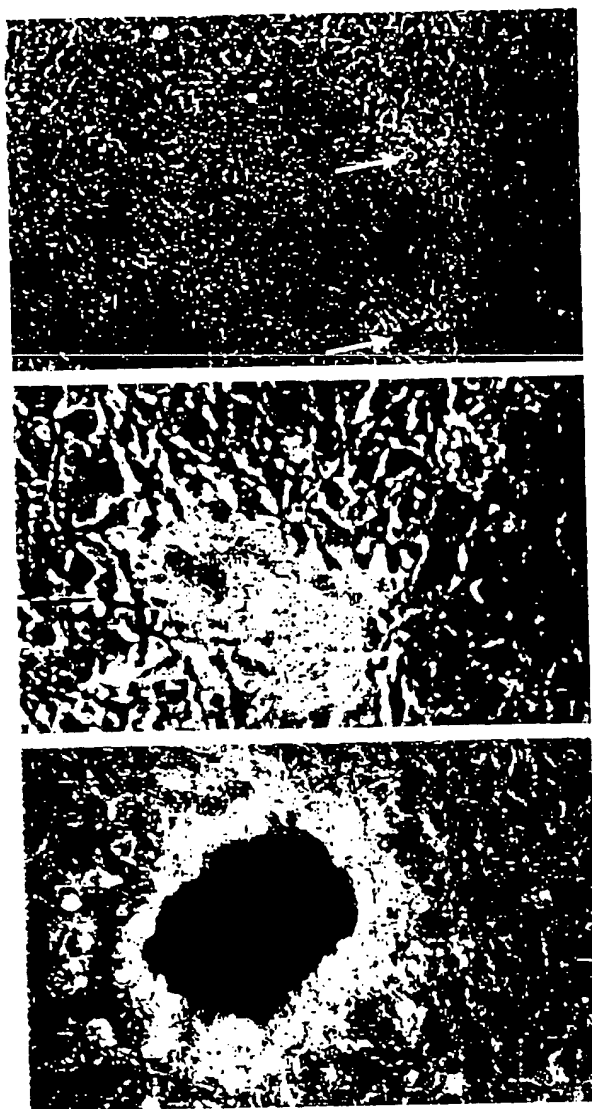


Figure 24 : Phase contrast images of small, medium sized and large foci induced in a monolayer of NIH-3T3 cells by transfection with pCB201.

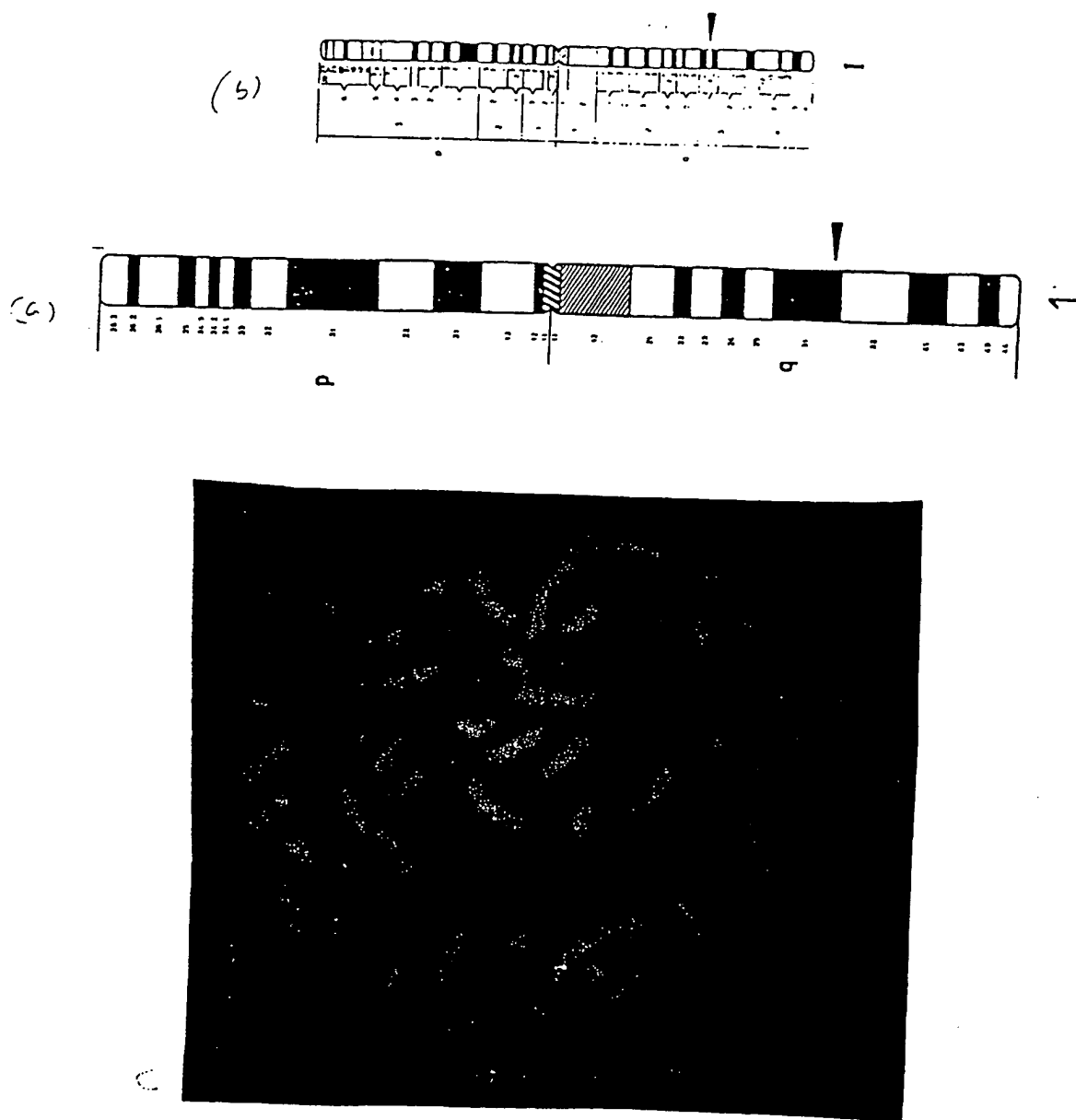


Figure 25a, b, c: Chromosomal localisation of hu-unc-53/1 by FISH

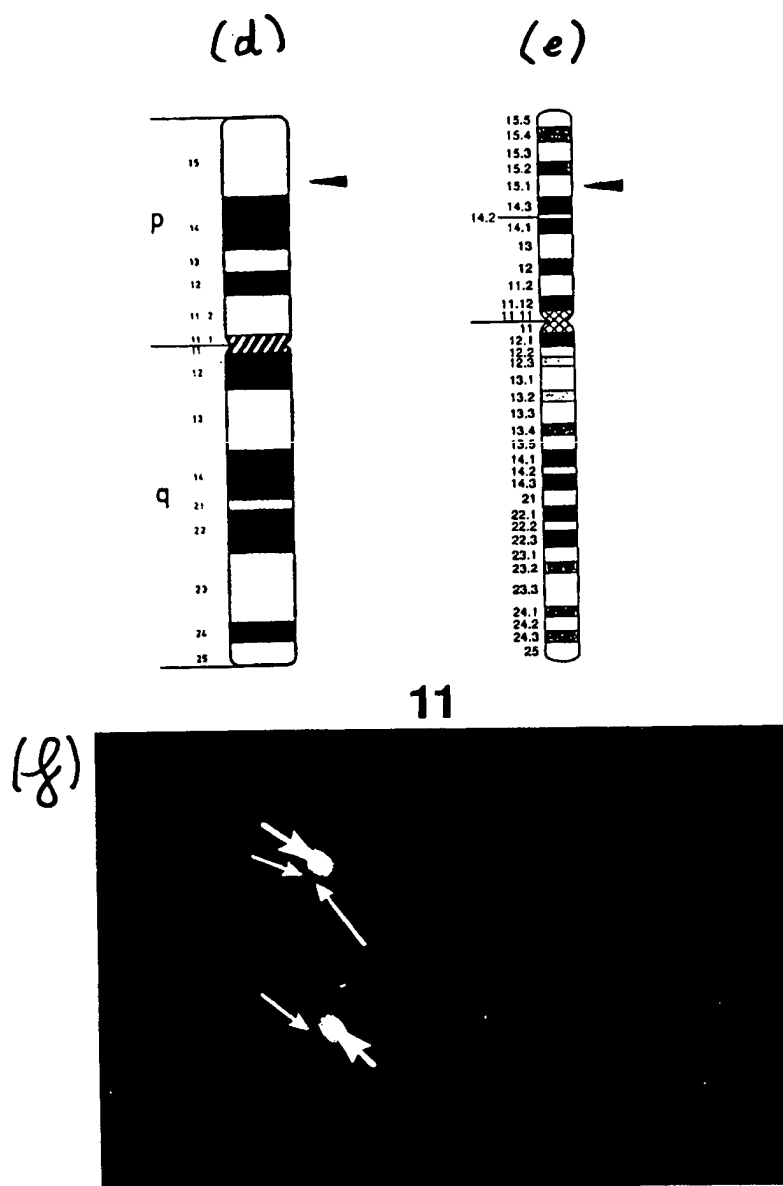


Figure 25d, e, f: Chromosomal localisation of hu-unc-53/2 by FISH

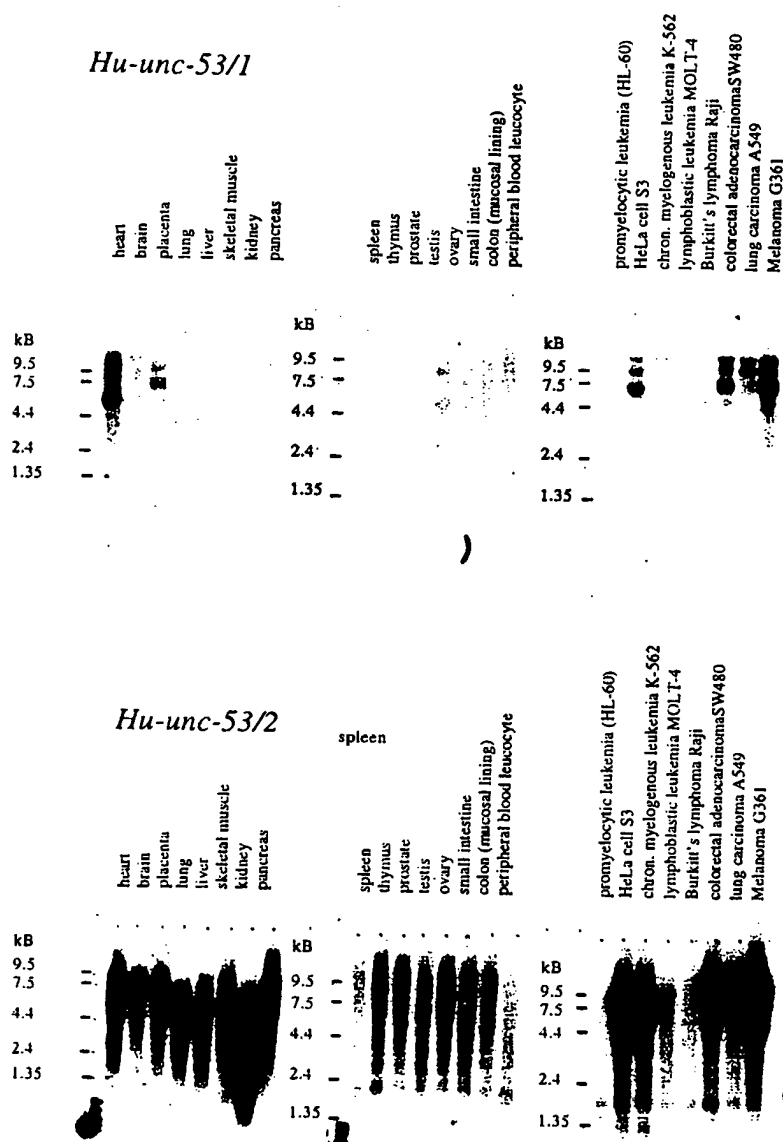


Figure 26 : Expression pattern of *Hu-unc-53/1* and *Hu-unc-53/2* in normal human tissues and cancer cell lines by Northern blotting.

donderdag, 27 november 1997 16:52

Page 1

fig 27 pNP3 Map (1 > 13621) Site and Sequence

Enzymes: All 146 enzymes (No Filter)

Enzymes: All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

[illegible]

donderdag, 27 november 1997 16:52
fig 27 PNP3 Map (1 > 13621) Size and Sequence

Page 2

ATTTTGGCCACATCGCAACACCGGCTCAGAGGTGTAGCGCTGGATTCTGAAATATTTTGAGAGTTATGAGGATTCAACTGTGGAAGGTTTTTTTCA 3100
AATTCTAACCCAATGTGCTCAAAAATTACTCTCGAAATTAATAATCTACAAGTGTTCATTCACCGAATTAGTGATTAACGTTCAAAATGTAGACAACAA 3200
GGAATAGGTAAGCATGTAGGCAATGTAAAGGTGTGGTAGGTCCAATAAACAATAATTATTAGCTTTTGATTATTCAGTGTTTTTTTAAATATTATT 3300
TACTGTTTATTGGCTTGGTAAATCAAGCCTATCTGCTCATCATACCTTTCTGCTACACACCTGCTATGCTAAAACTTAATACATCTCAAGAGA 3400
ACTTTCACAAATTCAAAAATTAATACTATAGATTCAAAAATTAACAAAAGAAAAGTTTTGGTACGAGCTCAACGTGTGTTAACTTTAAACGAAAC 3500
AGTTCAGAAACCAAAAATCAAAAGTTTTCTTTGACAAACATCATTTTTCTGGGATTTTTTAAATTTATAATTTGAAGCTATTTTCAACAAAAAAA 3600
TCAAAGTTTCAAAAATTTACCTGATGTCTTATCTTCCATCTCTCTCCAGCGAATGATTACCAGAAGTTTTGTGCGCGGAGTCACCTGAGCCGAC 3700
GCATGAGCCACAATGGAGTCTTCAGACGGGAGCTGCTCATATCTCGAAAATGGTCATTTGTCTATAATAATTGGAGGTGGGGGTGCACTGACCAAA 3800
CGACTTGACAGGCGATGTGATTTTTCTTGTTTAGAAACATCTGACGATAAGGTATTTGTGTGCGGTTTTTCTATGATGTTTCAAAAAAATTTA 3900
TCAGAGTAAAAAACAATAAGAGGTGTACACAGAACATTTGTGTAGATTTTGGGATAGAATTATTATTAATGGGGGAGGAGCACAATTTTTTCCACT 4000
TTTTATTACGAATTAATTTCTATTTTTGAATCTACACAATTTTATTGTAGAGGTGACTTAAAGAGCTAAAAACAAGAAATTAAGAGAGGTTTTT 4100
TGATAATAACAAAATAGGGGGTGGGAGCAGGGGAGAAAGAGCATGTAGGATGTTGAAGAGTAATTGGACTTAGAAGTCAGAGGACGGCGCGAGAT 4200
GCTGAAAATTTAAATATTACTGTTAGTTAGTAGAAGTCAAGTTAACTTACTCTGATGACAGCGGCCCTATTATTTTGACACCAGACAAGTTGGTAA 4300
TGGTAGCGACCGCGCTCAGTTGGAATTTCTACGAATGCTATTTGTATAGTTCACTCATGCCATGTGTAATCCAGCAGCTGTTACAACTCAAGAAGCAC 4400
CATGTGCTCTCTTTTTGTTGGGATCTTTCGAAAGGGCAGATTGTGTGGACAGGTAAAGTTGTCTGGTAAAGGACAGGGCCATCGCCAATTGGAGTA 4500
TTTTGTTGATAATGGTCTGCTAGTTGAACGCTTCCATCTTCAATGTTGTCTAATTTTGAAGTTCTGAAAATTTAAATATAATCCGTTAGTTAGTAA 4600
CCAAGTTTAACTTACAACTTGATTCCATCTTTTTGTTGTCTGCCATGATGTATACATTTGTGTGAGTTATAGTTGATTTCCAATTTGTGTCCAAGAA 4700
GTTTCCATCTCTTTAAATCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCACCTTCAAACTTGACTTCAGCACCTGAAAATTTAAATATGTAT 4800
GGTAGTTAGTACCGAAGTGTTTAACTTACGTGCTTGTAGTCTCCGTCATCTTTGAAAATATAGTTCTTCTCTGACATAACCTCGGGCATGGCAC 4900
TCTTGAAAAGTCATGCCGTTTCATATGATCTGGGTATCTCGAGAGCATTGAACACCATAACAGAAAGTAGTGACAAGTGTGGCTGAAAATTTAAATA 5000
ATCAGGGTTAGTTAGTATATATGTTTAACTTACCCATGGAACAGGTAGTTTTCCAGTAGTCAAAATAAATTTAAGGGTAAGTTTTCCGTATGTTGCA 5100
TCACCTTCACTCTCACTGACAGAAAATTTGTGCCATTAAACATCACCATCTAATTCACAAGAAATGGGACAACTCAGTGAAAAGTCTTCTCTT 5200
TACTCATTTTTTACCCGGTACCCTCAAGCGTACGTTGGTGGTGTGCGATTGAAGTCATTTTAAAGTGAACAACTGATTTTAGAGGTGAAGCGGTGG 5300
AAGTCCGTCTGTGATGCTCATTGATTGGAGCTTTTCAAGTCTGGTTTTTTCAGATATCAGGGGTCTGAAATTATAAAATATGAATTTGAGATTGAG 5400
GAGCATCTTCAAAAATACAAATATTTAAGATTCTAGTTTGAACAAAATTTATAGCCCAAGTCTGAAAATCGCCGAAAAGTGGAAAATGCTTGGT 5500
TTCAGAAATCACACATTTAAATTTTAAATTTTCTCAAAGGAATTTTTTCAATTTCTGCAATTTTCTCAATTTTGTTTTACAATTTTGGAAAT 5600
TTTTCTCAATTTTTCGGAAAATAGTGAATTTTGTATAGTTTTCCTATGTCTCTATTTCTGAAAATATGAATCTTTTGATTTTTTTTTTTGTTGAT 5700
TTTTTGAAAAGTAGCTTTATTTGTGAGTTTCAAAATATTTTTTGGAAATTAATTTCCGTTTTCTGATTTTTTGTGCTCTCGGCAACGAAAAATC 5800
GAATAAATTAATATGATTTTAAATATGTTTTGTAATTCACAAACAAATCTTACTATTAGACATATTTGGAAAAACATTTTGAAAAAATTTCCCAAG 5900
AATTTAGAAATTTTATAGAGTATTTTATTTTTTAACTCAATTTTTTAAACAACCTTCAGAAAAAACTTATGGTCAGAAAAATCAAGTTTACT 6000

donderdag, 27 november 1997 18:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence

Page 3

GAACCTTCTACAAATAAAATAATTGTACTCTCTTACTATYCCCCGACAACTTTAAATAATCCAAAAATTACTTCCATCAGATGGTCAACAAAAACG 6100
TCACCGACCGAAACATAAGTAAAGTTCAGAAATAAACGTACCATTTGTCAACTCTTATGTGCATAATTGGACTAACCACTCCGATTGTTGGCTGGGTG 6200
TCACGTGGCGGAACAGCTGGAGGTGGATCTTTTTTCGTGTGCTTTTCACTCCTTTCACTGCCAGTGTGGTTTTCTCTAAAAATTGATTATTTTTCGAA 6300
AATATAGGAAGGCTCTTAGGAAGCGAAGCCAGACATTAAAGTGTATCAAAATCTTAAATCTAAAAAGGGAATCTGCAACAAAGACGACAAAAA 6400
GGAAGAATGGAAAGACGGAAAGCGTATTGAATTCAGATATTTCCATTTTGCATTTTGCCTAGTCCGACTTTGAAGCCAGCTTCGAATCTTTAGAT 6500
TGCAACTACTCGGGCAAAGAATATGTACATGGTTCAGGCGGCGCCAGTTTAGCCATGCTTGATCAATGGCTTTGCCAGCTCTAAACCAAGTTAATAT 6600
AATTTAGGCCAAACTTGACAAGAAGCTCTACCCAGAAACAAAAAATACTCACCTCAACAGGCTTGTGATAATTTGGAGAAACCGGTGTTGGCGCTA 6700
TCGGCTGTCTGATAGCTGTACGATGGAGCGTTAAGAGTAAGATCGTCTGATGACGGAGACTTTTCGTCTGACGTTGAACCTGAAAATTTTGAAGTTT 6800
TATAGAAGGGGACGGTTTCAATGGGATAAATTGAACTTAAAGTTTCATCTCTAAAAATTCAGTTGCACAAAAATTAGTTTGAACCTTATAAATGAG 6900
ATTATTAAATCGAAGCTTCTCGGCGACACAAAAAGTTTAGGGACGGAACTCTATCTGAATGTTCTAAAAATTTATCCATATAAAAAAGTTTAAAGTA 7000
AAACATTTTTCTGAAATATCTTAACTTTTAACTCAATTTTGTATCTGAAAATAAAATTTAAAAAGTTTGGGTTTGACAAATACCGTATGACATTTAT 7100
CCTAAAAATGTATAAGCAGGTCTCTGCAAGTTTTTGTACCAACAAATGGGCGGATTCACATTTTATGGCAGGTGTCCAGAGACCTCACTTTTATA 7200
AACATCCGAGACTATGTGAAGTCAAGAAAAACAAAAAACAGTTCTCTAAACACGAACCTTGAAGATGTGAATGCATGCTTAGGGAACCTTCG 7300
TCGATGATGACGTTGGCGACGTGCTGTTGAATCCAGCGTATCCGGACTCTTCTCACTGCTCTTTGAGCATCGTTTCAAGTCTTGTGAGATATGATTGG 7400
GGCGTCCGTTTACGGTAGGAACTTTTGGCTGGAAATTACTATGTAAAAACGAAGATGCACAGTTTTTTTTCAAAAAATTTTAAAAAATGTTTTT 7500
TCATAAATAATAGAAGCCATAAGACAATTCGGAAAGTCAAAAGATCTCGTATATTTCACTAGGAAGTTTTAAAAAATATAGCAGGTCTAAGAGCTCT 7600
AATTTATATGAAAAAATAAATACTAGAACTGAAACGTAAGTTAATTTTCAAAAGTGCATCTCACAAAAATGTTTGAATTTGTGCTCTTTC 7700
ACTCTGTAGTAGCTCATTTTGAACAAAGTGAACAAACCGGCGAAGAGAGCTGAGAAATTGCTTTACTCTTTGTGACGCAGATGTGTTTTCGGCCCT 7800
CGACGTGAAAAATGCTCGGAATATCACTTCCGTTGCAACATTTTCCGCAAGAGCTTTTATGGGGGACACAAAGCAAAAAAGATGGGACATAGATTC 7900
TGAAGAGTGTCAAGGAGAGAGAGAAAAATGGGAGACAATTCCEAATGTCTCATCATTTTATTAGTAGGTGGTTTTGTTTTGGTAGAGTCAAGGAA 8000
CGGAAGAAGTACGAATCGGGTTCTAAATATAGATGAGGCCAGAGCTTCCGCAAACTAGGGTTTTAAAAATTTCTGAAAATCAACTTTGACAGCTTATAT 8100
CTCACTGTCTTTGGTCTATACAAAAATTATCACTGTCAAAATATTTTTTATTTCTTTTCTATATCTGATATCTCAATATGAACCAAGATATAAAT 8200
CTTCAAGTAAATAGTGGGATGCAATACTATCAGAGGAAAAATCTTATTTTAAATTTAATTTGTAGAAATTTCAAGTTTAAAAAATACATTTCCAAT 8300
TTCCGATTTCAAACTACTACCGTACAAAGCTTCGACATAGACGTGGCACTTCCAGCTTACTGCTCGGCGGCTTCAAGGCACTTTCACTGGGGCAGCGA 8400
TTTTGACAAAGTTTGTGTTGAGGCACCGCCGCCCTTTCTCGTAGGTGTGGGTATTGATGAGGAAGATGGGTTTTGCTACTGAATAACTTTAA 8500
TTTTGAGATTCCACCACCACCACCGCTGTATCGGCTCTGTTTTGCTCCAATAGTCTTCACAGAAGCAAGTTTGGGGTGTCCAGGCTTTCGGA 8600
GCGGTAGCTTTGAGCTTCGATTTTTGTAGTTGTAGCAACACGAAGTCTGGGTTTGTGGTCTAGAGGTTTTTGGAGTTGGGAGGTAGGTGCGTTA 8700
GATTCGAAATAGAGCTGTACGTTGATGATGATTCTGAAATAATATAATTTCTAAAAAGAAGCCGGAAGGATCGGATACCTAAGCTTCTCGAGATGTGGA 8800
TATCTCGAGCCAACTATTATTTGCCACTCGAACGGCTCGACGGACGGAATGAATTTGTATTTGATGAAGTGGTTGATGATGAGGGTGGTTAAGT 8900
CCAGAGCTCTTTGGCTTGATACCAATCTTTGATGAATCTGTAATTGAAAGTGTGACGAGAAAGGATAACGATTTTACCGCATCAGGGCATTAAAGATTT 9000

Page 4

BNSDOCID: <WO__9824810A2_I_>

donderdag, 27 november 1997 16:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence

Page 5

GTTCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGGTGT 12100
AGGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTTCAGCCCGACCGCTGCGCTTATCCGGTAACATCGTCTTGAGTCAACCCGGTAAG 12200
ACACGACTTATCGCCACTGGCAGGCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTTCTTGAAGTGGTGGCTAACTAC 12300
GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCG 12400
CTGGTAGCGGTGGTTTTTTTGTTCGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA 12500
GTGGAAACGAAACTCAGCTTAAGGATTTTGGTCATGAGATTATCAAAAAGGATCTCACTAGATCCTTTTAAATTAATAATGAAGTTTAAATCAATC 12600
TAAAGTATATAGTAACCTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTGGTTTCATCCATAGTTGCTT 12700
GACTCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATCATACCGGAGACCCACGCTCACCGGCTCCAGATTT 12800
ATCAGCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCACTTTATCCGCTCCATCCAGTCTATTAAATTGTTCCCGGGAAGCTAGA 12900
GTAAGTAGTTCGCCAGTTAATAGTTTGGCAACGTTGTTGCCATTGCTACAGGCATCGTGCTGTACGCTCGTCTTTGGTATGGCTTCATTCAGCTCCG 13000
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCTTCGGTCTCCGATCGTTGTGCAAGTAAGTTGGCCGC 13100
AGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCTGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA 13200
TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGGCGTCAATACGGGATAATACCGGCCACATAGCAGAACTTTAAAGTGCTCATATTG 13300
GAAAACTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTGTAGATCCAGTTCGATGTAAACCACTCGTGACCCCACTGATCTTCAGCATCTTT 13400
TACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC 13500
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA 13600
CATTTCCCCGAAAAGTGCCAC 13621

FIGURE 28.

SIGNATURE SEQUENCES :

Different signatures can be used to define to identify the UNC-53 gene family :

Aminoacids are listed in one letter code
 X equals any aminoacid
 X(3,5) equals 3 to 5 X's
 (D,E) means D or E at a given position

The signatures should be used to screen a database using a weight matrix of conservative substitutions.

BLOCK A :

GSXLSLASX(3,5)YXSXXEX(4,5)IRXXXR(D,E)LEXXXVXXLTXXXXXXXLX
 XXFEQXL

BLOCK B :

KXKKSXXXXXXXFXK

BLOCK C :

LARGE FAMILY :

VXXL(K,R)XELX(D,E)(R,K)(D,E)XXLXX(V,I)RL(D,E)XLXXAXXXDXLRE(T,A)X
 XXXXXEXXXLKXEXD(R,K)LX

VERTEBRATE FAMILY :

VXXLRXELX(D,E)(R,K)(D,E)MKLTDIRLEALXSAHQDQLLREXMXXMQXEXX
 LKAENDRLK

BLOCK D :

LARGE FAMILY :

W(K,D)X(I,L)DXX(I,V)XX(L,V)F(K,E)XY(I,V,L)XXXDXXXLXG(I,L)X(2,3)(D,E)S
 (I,V)XGYXI(G,S)(E,H)(L,V,I)(R,K)(R,K)

VERTEBRATE FAMILY :

FXXGCXXVSGKXXWXXLDXXVXX(L,V)FK(D,E)YIXXXDPXXXLG(I,L)XX(D,E)
 S(I,V)XGYXI(G,S)XXKR

BLOCK E :

GXXGXGKS/T

and

F(K,R)MXXXSXX(3,8)GF(I,L,V)(I,L,V)(R,K)Y(I,L,V)(R,K)(R,K)(R,K)XV(D,E)

and

F(I,L)EKXXXX(D,E)XXXGPXXX(L,I)XCPXXXXXX(R,K)XWFXLWNXXXIPY(
 L,I)XXXA(R,K)DGX(K,R)XXGXXXX(F,W)EDP

Block F

(W/F) (D/E) DSSS (V/L/I) SSGISD (T/N)

Tuesday, 18 November 1997 10:34

fig 29 pEGFPsac (1 > 5100) Site and Sequence

Enzymes: 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

7p.

100 TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTACGGTAAATGSCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTACCAGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

200 CCAACGACCCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGSTTGCTGGGGGCGGTAACCTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACCTGCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

300 AAACGCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGSCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTACCAGGGCGGACCGTAATACGGGTCTAT
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

400 CATGACCTTATGGGACTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W

500 TAGCGGTTTGACTCACGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTGTTTTGGCACCAAATCAACGGGACTTTCCAAATGTCGTA
ATCGCCAACTGAGTGCCCTAAAGGTTAGAGGTGGGGTAACCTGCAGTTACCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S

600 ACAACCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGTAACCTGCGTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTGCTCTGACCAAATCACTTGGCAGTCTAGGCGATCGCGAT
O L R P I D A N G R . A C T V G G L Y K Q S W F S E P S D P L A L

700 CCGGTGCCACCATGGTGAGCAAGGCGGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAACGCGCCACAAGTTCAGCG
GGCCACGGTGGTACCACCTCGTTCCTCAGCAAGTGGCCCCACCAAGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTCAAGTCCG
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

800 TGTCCGGGAGGGCGAGGGCGATGCCACCTACGSCAAGCTGACCTGAAGTTCATCTGCACCAACGGCAAGCTGCCCGTGCCCTGGGCCACCTCGTGAC
ACAGGCGCGTCCCGCTCCCGCTACGGTGGATGCCGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCACGGGACCGGGTGGGAGCACTG
V S G E G E G D A T Y S K L T L K F I C T T G K L P V P W P T L V T

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 2

CACCCGTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAASCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCTGTCTGAAGAAGTTACGGCGGTACGGGCTTCGATGCAGGTCTCTC 900

T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

KspI

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGTCTCCGCTGTGGGACCACCTGGCGTAGCTCGACTTCCCGTAGC 1000

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTTCTTGGCGTAGTT 1100

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACACGAGAACACCCCATCGGCGACGGCCCGTGCTGCTG
CCACTTGAAGTTCAGGCGGTGTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200

V N F K I R H N I E D G S V O L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCACATGGTCTGCTGGAGTTCGTGACCSCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTACGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGATACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300

P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G

BspM II Bgl II

Asu II
EcoN I

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCAATCGGCACCTTCT
AGTGAGAGCCGTACCTGCTCGACATGTTACGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCCTAACCCGTTAGCCGTGGAAAG 1400

C.e.unc53 sac

I T L G M D E L Y K S G L R S T S H V E L I P I Y T O W A N R H L S

Nru I

EcoR I

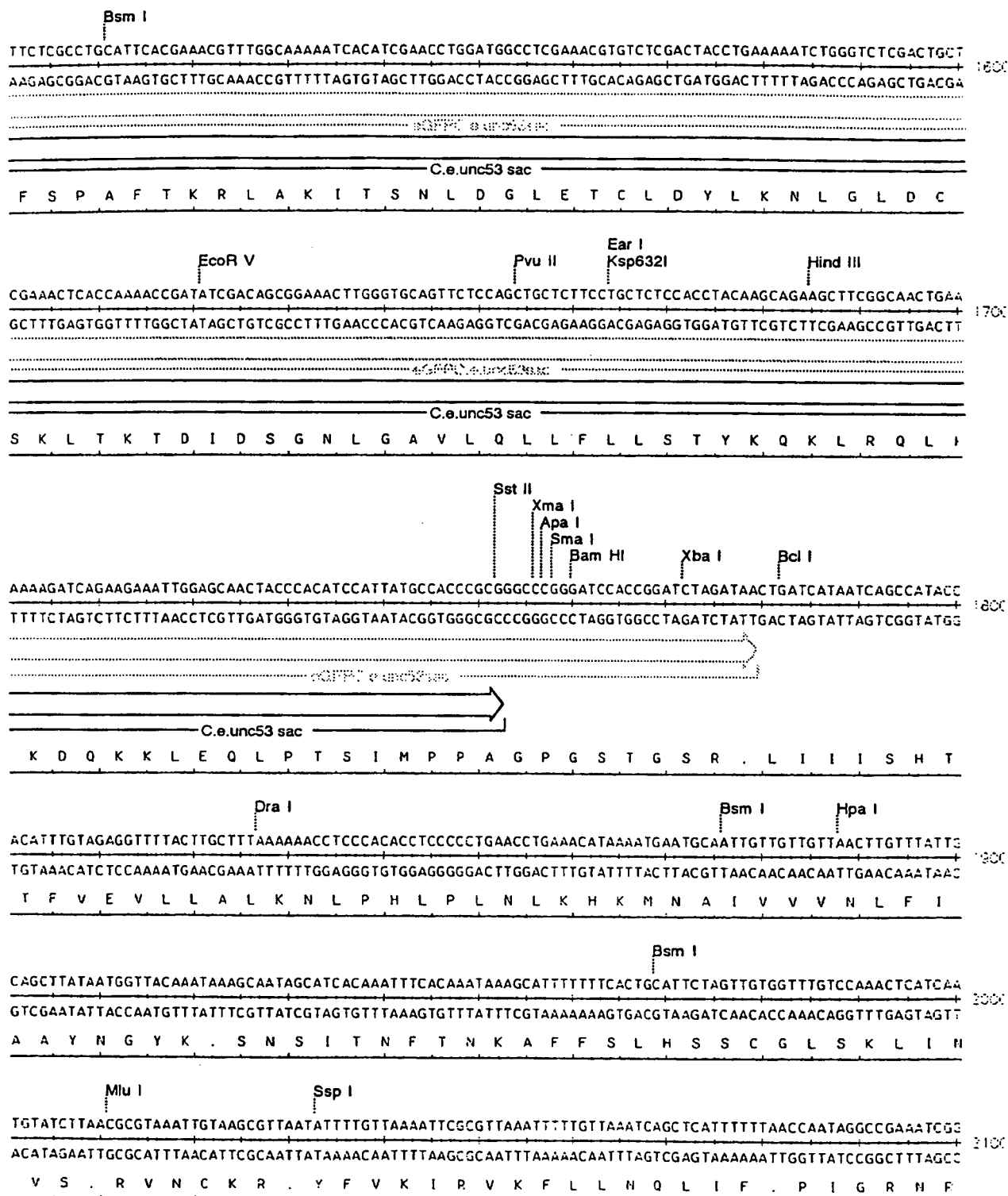
GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAGCGCTATAGCTGACCAAGAGTTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500

C.e.unc53 sac

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 3



Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 4

Bsr I

CAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCAGTTTGGAAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTC
GTTTTAGGGAATATTTAGTTTCTTATCTGGCTCTATCCCACTCACAACAAGGTCAAACCTTGTTCAGGTGATAATTTCTTGACCTGAGGTTSCAG 1200
Q N P L . I K R I D R D R V E C C S S L E Q E S T I K E R G L Q R

Dra III

AAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGA
TTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTCAAAAAACCCAGCTCCACGGCATTTCTGTGATTTAGCCT 1300
Q R A K N R L S G R W P T T . T I T L I K F F G V E V P . S T K S E

Nae I

ACCCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCT
TGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCTTTCGGCCGCTTGACCGCTCTTTCCTTCCTTCTTCGCTTTCCTCGCCCGGATCCCGCGA 1400
P . R E P P I . S L T G K A G E R G E K G R E E S E R S G R . G A

Ksp I

GGCAAGTGATAGCGTACGCTGCGGTAACCAACACACCCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAATGTGCGC
CCGTTACATCGCCAGTGCAGCGCGCATTTGGTGGTGTGGGCGGCGCAATTACGCGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCTTTACACGCG 1500
G K C S G H A A R N H H T R R A . C A A T G R V R V H F S G K C A

BspH I

GGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
CCTTGGGGATAAACAAATAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTACGAAGTTATTATAACTTTTCTCTCT 1600
R N P Y L F I F L N T F K Y V S A H E T I T L I N A S I I L K K E E

Ssp I

Ear I

Ksp632I

OxaN I

Pvu II

Sph I

Ava III

Nsi I

GTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCAGGCTCCCCAGCAGGCAGAAAGTATGCAAGCATGCATCTC
CAGGACTCCGCTTTCTTGGTCGACACCTTACACACAGTCAATCCACACCTTTACAGGGTCCGAGGGGTCGTCCGCTTTCATACGTTTCGTACGTAGAG 1700
S . G G K N O L V N V C Q L G C G K S P G S P A G R S M Q S M H L

Sph I

Ava III

Nsi I

AATTAGTCAGCAACCAGGTGTGGAAGTCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
TTAATCAGTCGTTGGTCCACACCTTTACAGGGTCCGAGGGGTCGTCCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGCGCGG 1800
N . S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P

Bsr I

Nco I

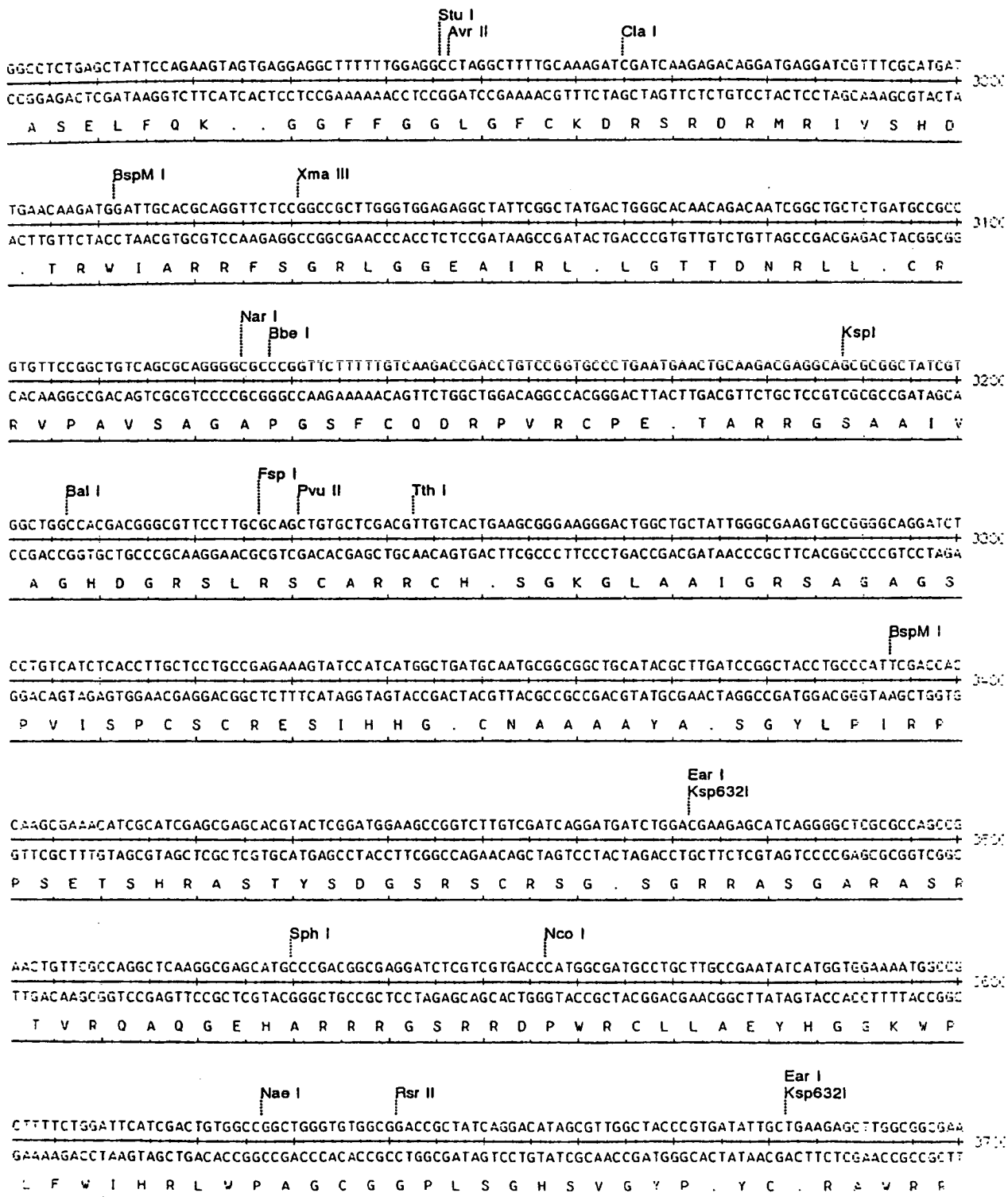
Bgl I

Sfi I

TAACTCCGCCATCCCGCCCTAACTCCGCCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTC
ATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGGTACCGACTGATTAATAAATAAATACGCTCCGGCTCCGGCGGAG 1900
L T P P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 5



Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 6

TGGGCTGACCGCTTCTCGTGCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCAGAAATGCCATAGCGGGGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA 3800
M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G T L

Asu II BspM I
GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTT
CCCCAAGCTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAG 3900
G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I V F

Nae I KspI Avr II
CGGGAGCGCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCTTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCCCTGCGGCGGACCTACTAGGAGGTGCGGCCCTTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATG 4000
R D A G W M I L O R G D L M L E F F A H P R G R L T E T R K E T I

KspI
CGGAAGGAACCGCGCTATGACGGCAATAAAAGACAGAATAAAACGCACGGTGTGGGTGCTTTGTCATAAACGCGGGGTTGGTCCCAGGGCTGGCA
GCCTTCTTGGGCGGATACTGCGGTTATTTTCTGTCTTATTTGCGTGCCACAACCCAGCAAAACAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGT 4100
P E G T R A M T A I K R O N K T H G V G S F V H K R G V R S Q G W H

CTCTGTCGATACCCACCGAGACCCCATTTGGGGCAATACGCCCGGTTTCTTCTTTTCCCCACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTC
GAGACAGCTATGGGGTGGCTCTGGGGTAACCCGCTTATGCGGGCGCAAGAAGGAAAGGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAG 4200
S V D T P P R P H W G Q Y A R V S S F S P P H P P S S G E G P G L

AlwI OxaN I Dra I Dra I
GCAGCCAACGTCGGGGCGGACGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGG
CGTCGGTTGCAGCCCCCGCTCGGGACGGTATCGAGTCCAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCC 4300
A A N V G A A G P A I A S G Y S Y I L . I D L K L H F . F K R I .

BspH I
TGAAGATCCTTTTGTAAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTT
ACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGSAATTGCACTCAAAAGCAAGGTGACTCGCAGTCGCGGCATCTTTCTAGTTTCTAGAGAAAC 4400
V K I L F D N L M T K I P . R E F S F H . A S D P V E K I K G S S .

AGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCACAAAAAACCACCGCTACCAGCGGTGGTTGTTTGGCGGATCAAGAGCTACCAACTCTTTT
TCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTGGTGGCGATGGTTCGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 4500
D P F F L R V I C C L O T K K P P L P A V V C L P D Q E L P T L F

BsrI
CCGAAGGTAAGTGGCTTCAGCAGAGCGCAGATACCAAACTGTCTTCTAGTGTAGCGGTAGTTAGGCCACCACCTTCAAGAACTCTGTAGCACCAGCTA
GCTTCCATTGACCGAAGTCGCTCGCGTCTATGGTTTATGACAGGAAGATCAGATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATGTCGCGGAT 4600
P K V T G F S R A Q I P N T V L L V . P . L G H H F K N S V A P P

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 7

AlwI I

CATACCTCGCTCTGCTAATCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCG
T Y L A L L I L L P V A A A S G D K S C L T G L D S R R . L P D K A

2700

ApaI I

GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
CGTCGCCAGCCGACTTGGCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGC TGGATGTGGCTTGACTCTATGGATGTGCGACTCGATACTCTTTCC
Q R S G . T G G S C T Q P S L E R T T Y T E L R Y L Q R E L . E S

2800

GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCCTGGT
CGGTGCGAAGGGCTTCCCTCTTCCGCTGTCCATAGGCCATTCGCCGTCCAGCCTTGCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTCGGGACCA
A T L P E G R K A D R Y P V S G R V G T G E R T R E L P G G N A W

2900

ATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
TAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGGTCGTTGCG
Y L Y S P V G F R H L . L E R R F L . C S S G G R S L W K N A S N A

3000

Ava III
Nsi I

GGCCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTGTTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
A F L R F L A F C V P F A H M F F P A L S P D S V D N R I T A M H

3100

Tuesday, 18 November 1997 10:34

fig 30 pEGFP72 (1 > 9697) Site and Sequence

Enzymes: 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Page 5

16p

100 TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGAC TGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K W P A V L T

200 CCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTGCTGGGGGGGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

300 AAACGCCCACCTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTTACATAGTATACGGTTTATGCGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTAT
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

400 CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCGCGACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W

500 TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAACTGAGTGCCCTAAAGGTTTCAAGGTTGGGGTAACGTCAGTTACCTCAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .

600 ACAAACCCGCCCATTTGACGCAAAATGGGCGTAGGCGTGACGGTGGGAGGTCATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACGCGTTTACCGCCATCCGCACATGCCACCTCCAGATATATTCGCTCGACCAAACTCACTTGGCAGTCTAGGCGATCGCGAT
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L

700 CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTCCCATCTGGTTCGAGCTGGACGGCGACGTAACGSGCCACAAGTTACGCG
GGCCAGCGGTGGTACCCTCGTTCCCGCTCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTGGCGGTGTTCAAGTCGC
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

800 TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGAC
ACAGGCGGCTCCCGCTCCCGCTACGGTGGATGCCGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCGGTCGACGGGCACGGGACGGGTGGGAGCACTG
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T

eGFP.C.e.unc53

eGFP.C.e.unc53

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 2

CACCC TGACCTACGGCGTGCAGTGCTTCAGCCGC TACCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCGAAGGCTAGTCCAGGAG 900
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCAGTACGGGCTTCCGATGCAGGTCC TC
-----eGFP.C.e.unc53-----
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
-----KspI-----
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG 1000
GCGTGGTAGAAGAAGTTCCTGCTGCCGTGTATGTTCTGGGCGGGCTCCACTTCAAGCTCCCGCTGTGGGACCACCTGGCGTAGCTCGACTTCCCGTAGC
-----eGFP.C.e.unc53-----
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCTCGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAGCGCATCAA 1100
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTCTTCCCGTAGTT
-----eGFP.C.e.unc53-----
D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

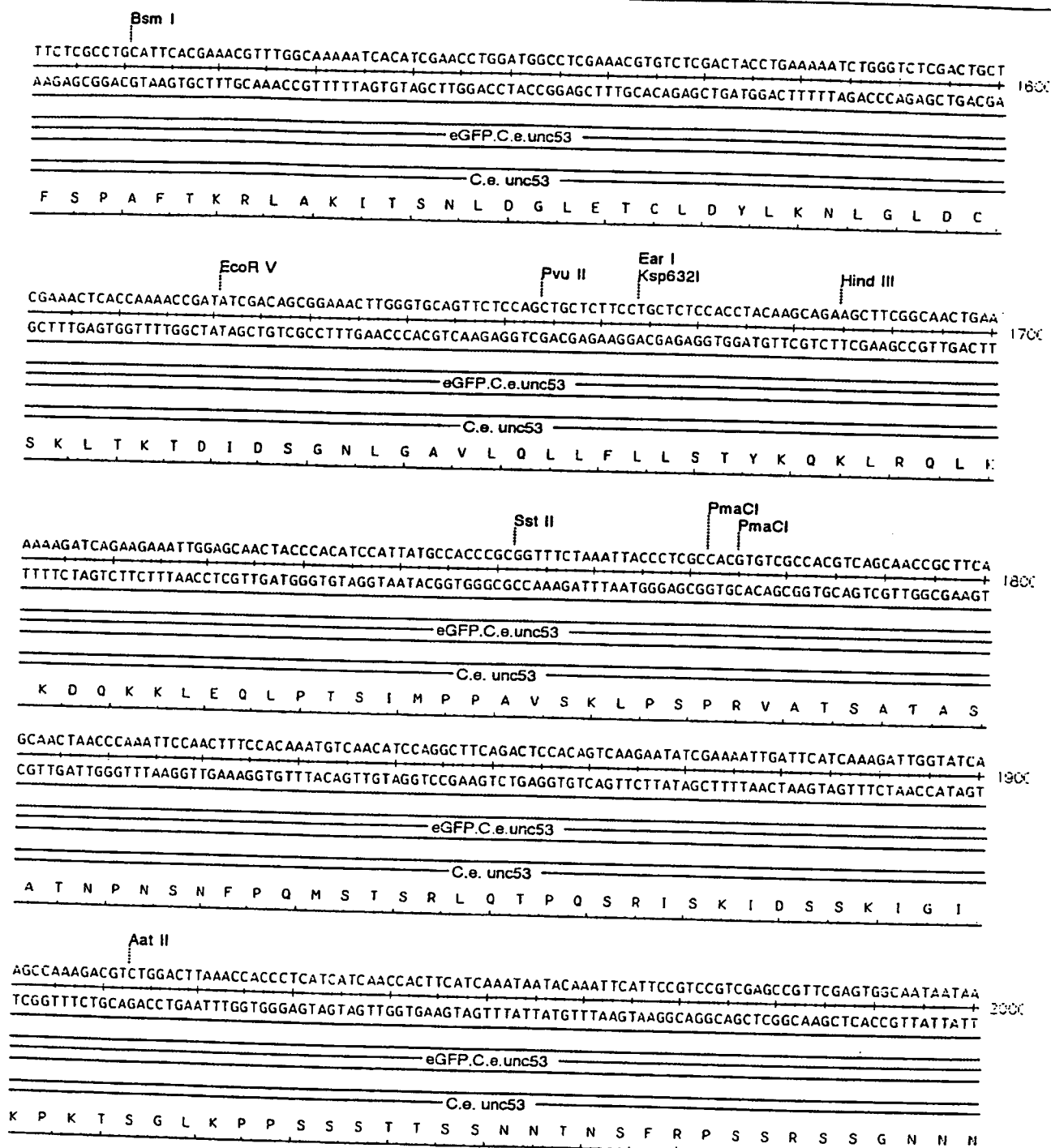
GGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGGCAGCGGCCCGTGCTGCTG 1200
CCACTTGAAGTTCAGGCGGTGTGTAGCTCCTGCGCTCGCACGTGAGCGGCTGGTGTGCTGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC
-----eGFP.C.e.unc53-----
V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGA 1300
GGGCTGTGGTGATGGACTCGTGGGTACGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGGCGCCCT
-----eGFP.C.e.unc53-----
P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

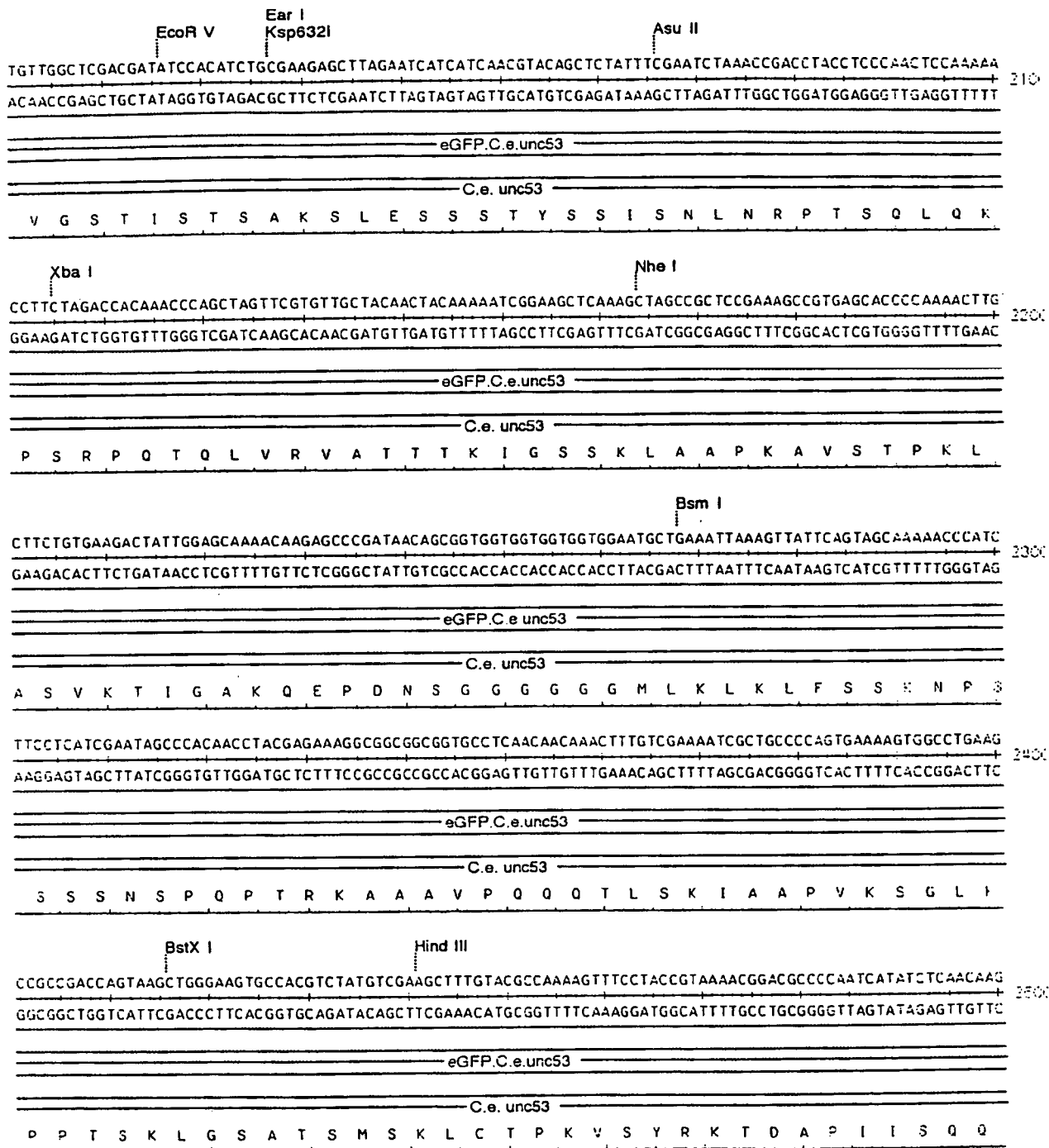
TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTT 1400
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGCCTGAGTCTAGATGCAGTTTACATCTTAACTATGGTTAGATGTGCTTAACCCGGTTAGCCGTGGAAAG
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S
-----Nru I-----EcoR I-----
GAAGGGCAGCTTATCAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA 1500
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 3

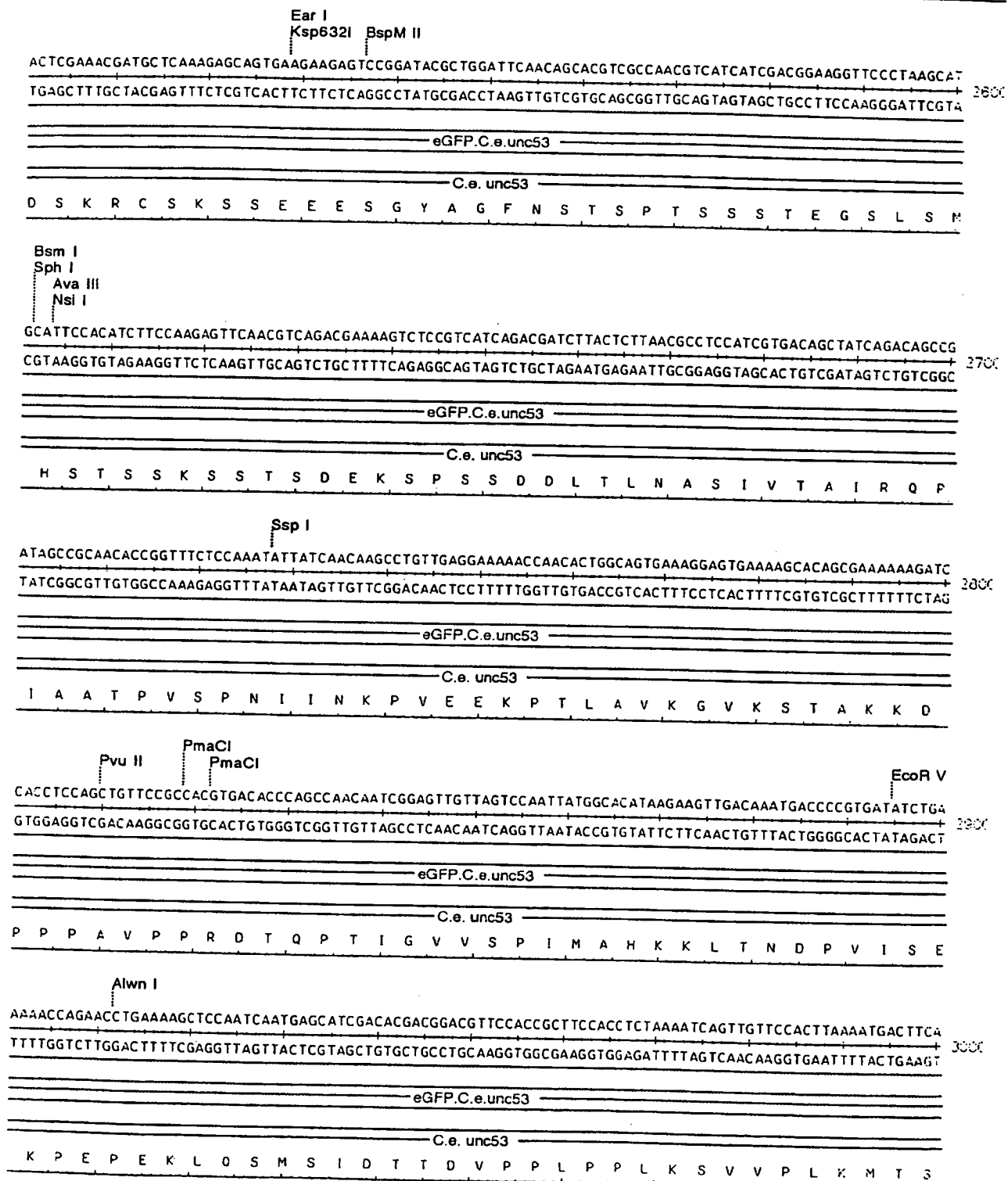


Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1>9697) Site and Sequence



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 5



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4

SpII
SpII

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCGCGCTCGAAGACT 3100
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTGTAGTGTAGCGGACAGTTCAGCAAACCTATACTCGTCAGCAGGCGCAGACTTCTGA

eGFP.C.e.unc53

C.e. unc53

I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC 3200
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGTCTTAG

eGFP.C.e.unc53

C.e. unc53

S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S

CAGCGGCTACACCTCTGACGCGGTTGCGATGTGCGCCAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT 3300
GTGCGCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGCTTTGCCGATA

eGFP.C.e.unc53

C.e. unc53

S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A O N G Y

Asu II Sst I BspM II

CCTGACAACCTCGAAGACAGTTCTCTCTTGTGCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTGTCCGGAGTAGACATGG 3400
GGACTGTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCTTAACAGGCCTCATCTGTACE

eGFP.C.e.unc53

C.e. unc53

P D N F E D S S S L S S G I S D N N E L D D I S T D D L S G V D M

CAACAGTCGCCTCCAAACATAGCGACTATTCCTTCTGTCGCTCCACGCTCTTCTCTCAAGCCCGAGTCCCAGTCGGTCTCCACATCAGT 3500
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA

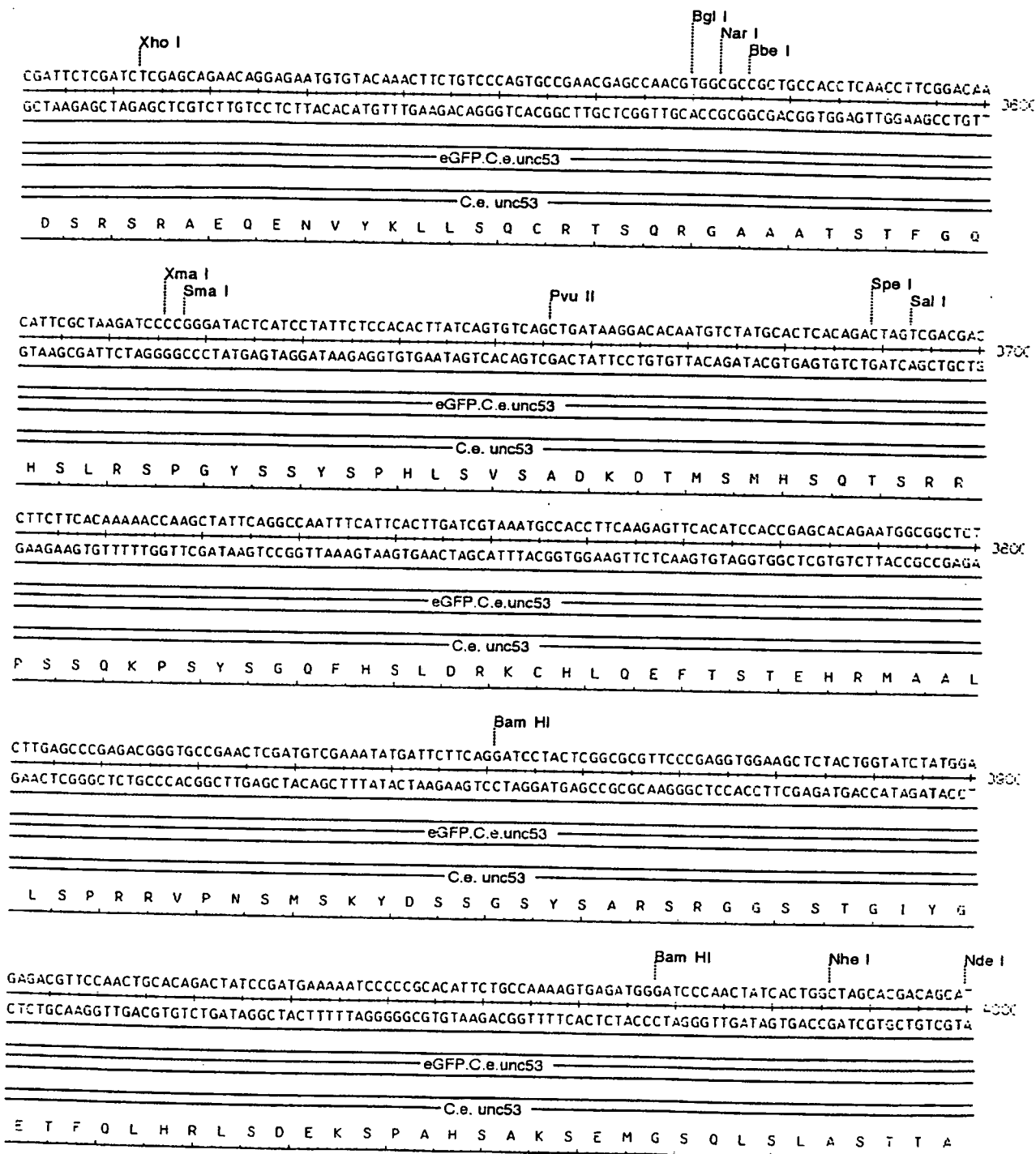
eGFP.C.e.unc53

C.e. unc53

A T V A S K H S D Y S H F V R H P T S S S S K P R V P S R S S T S V

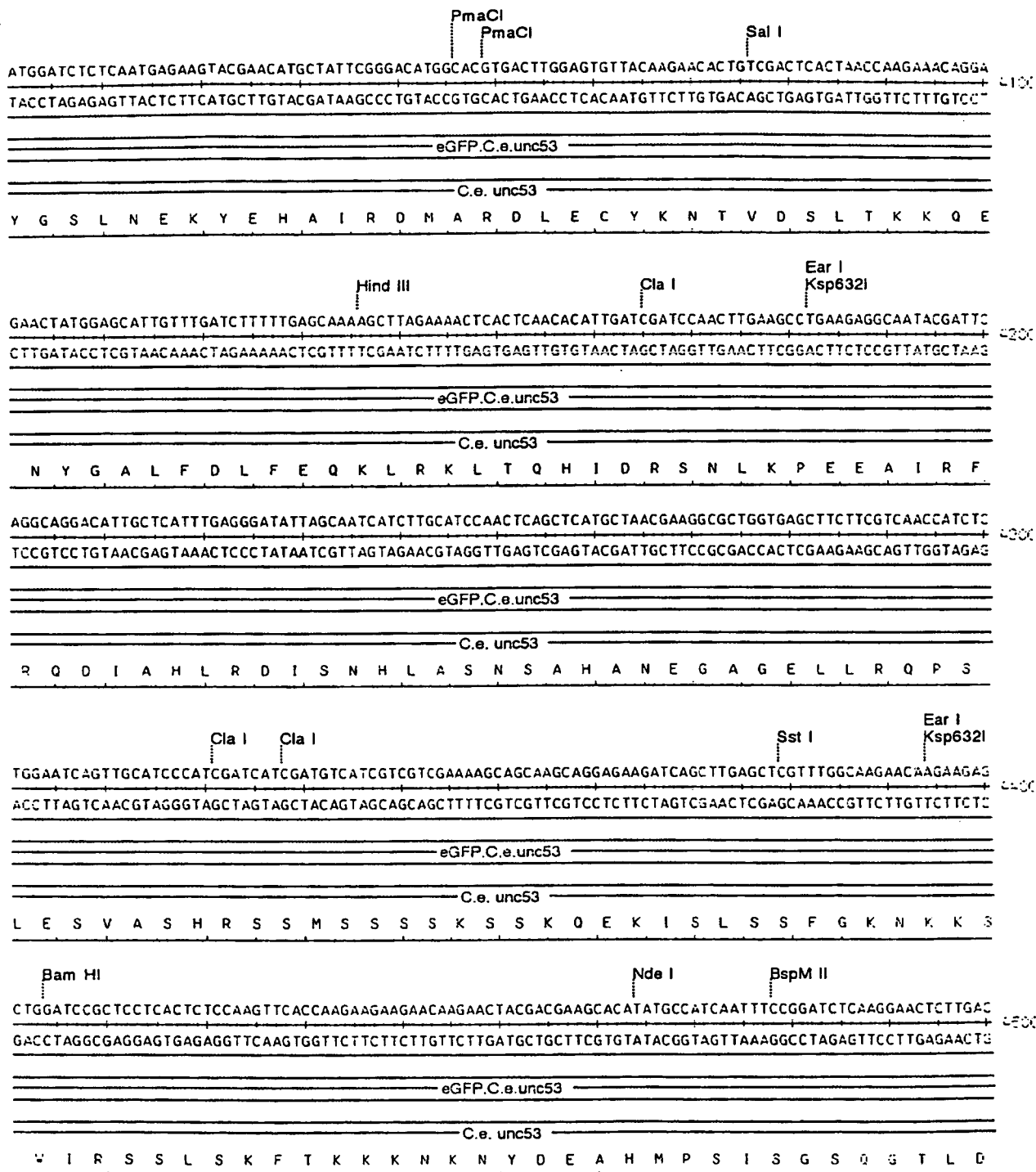
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 7



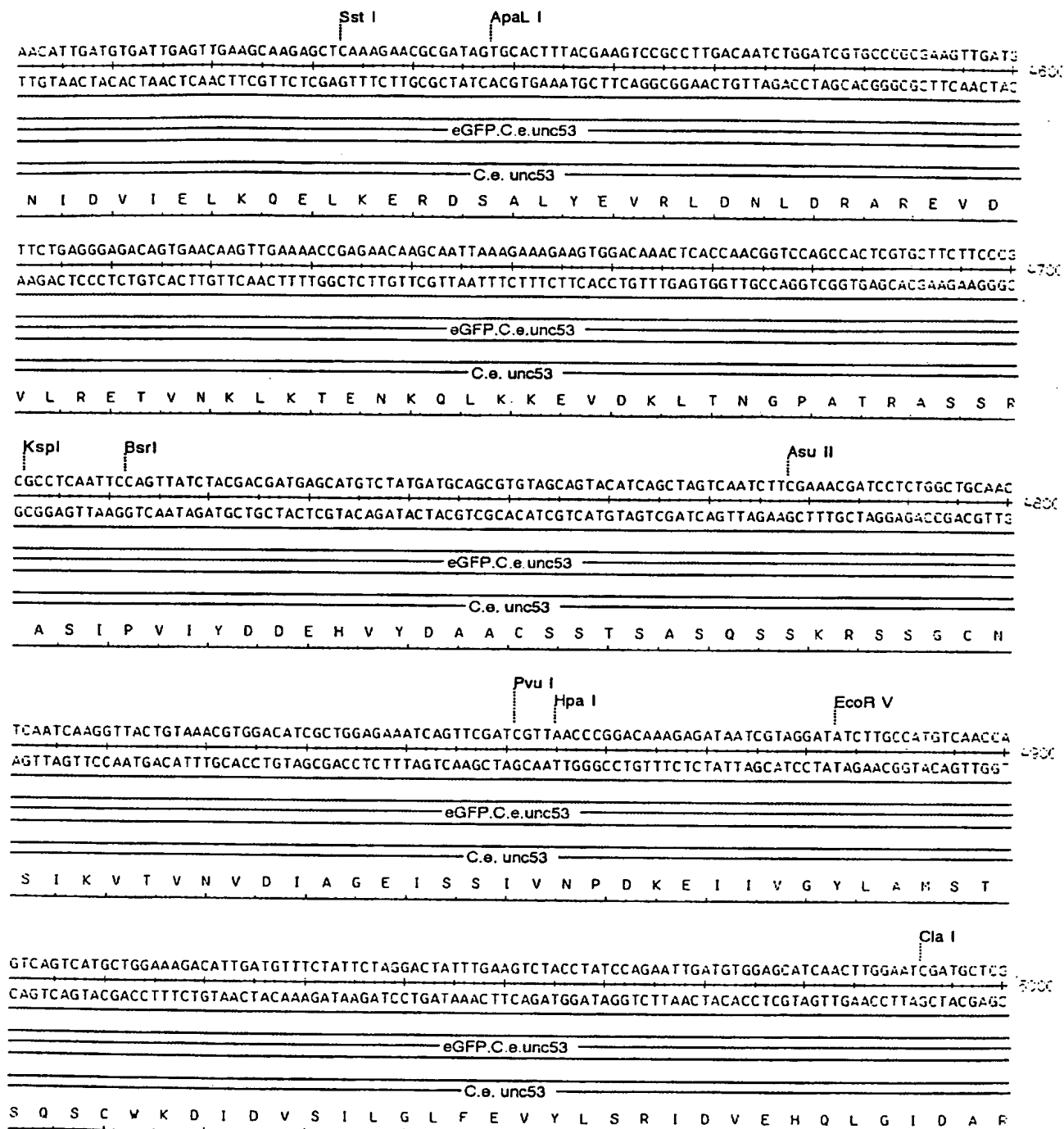
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 8



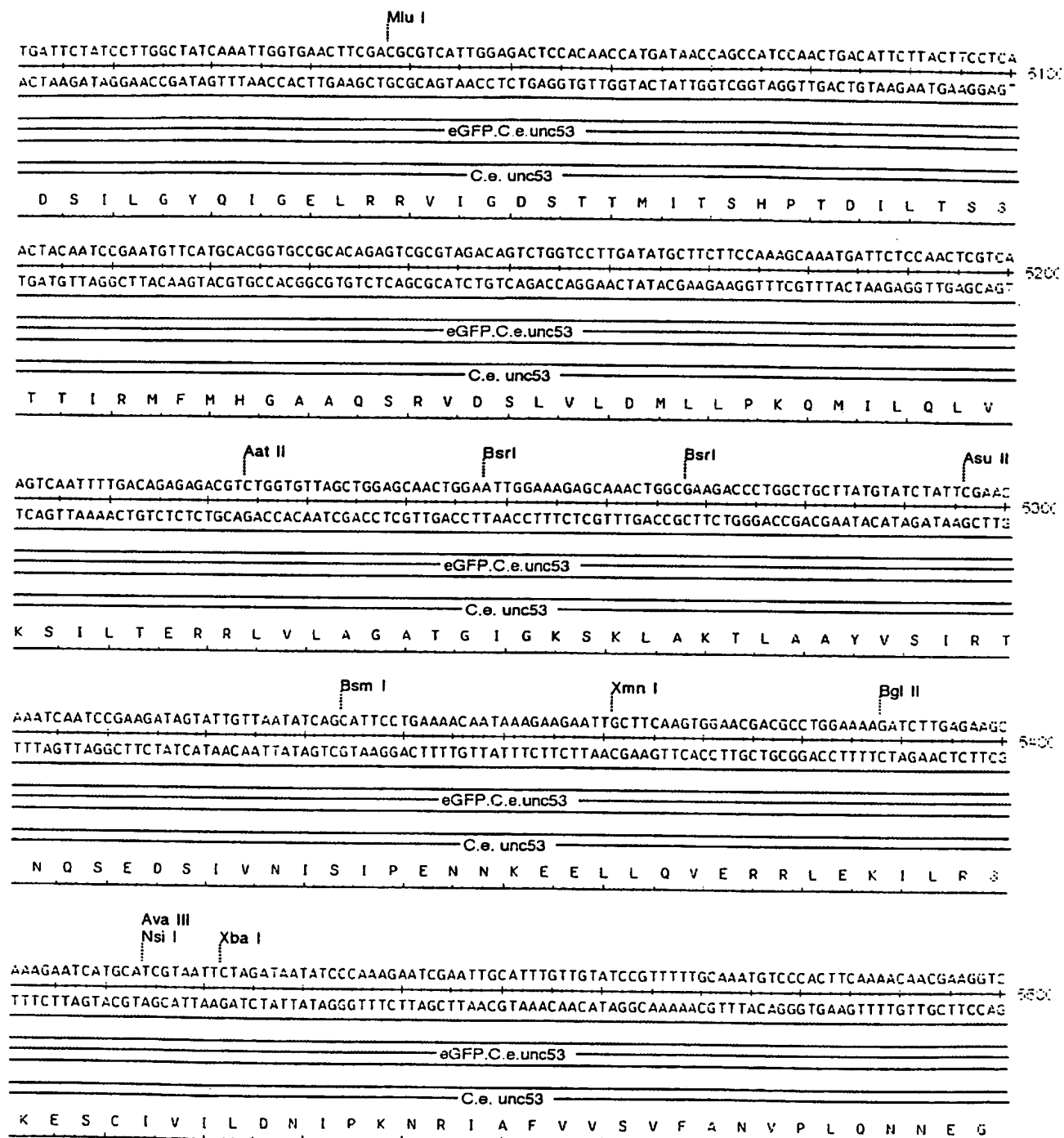
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 9



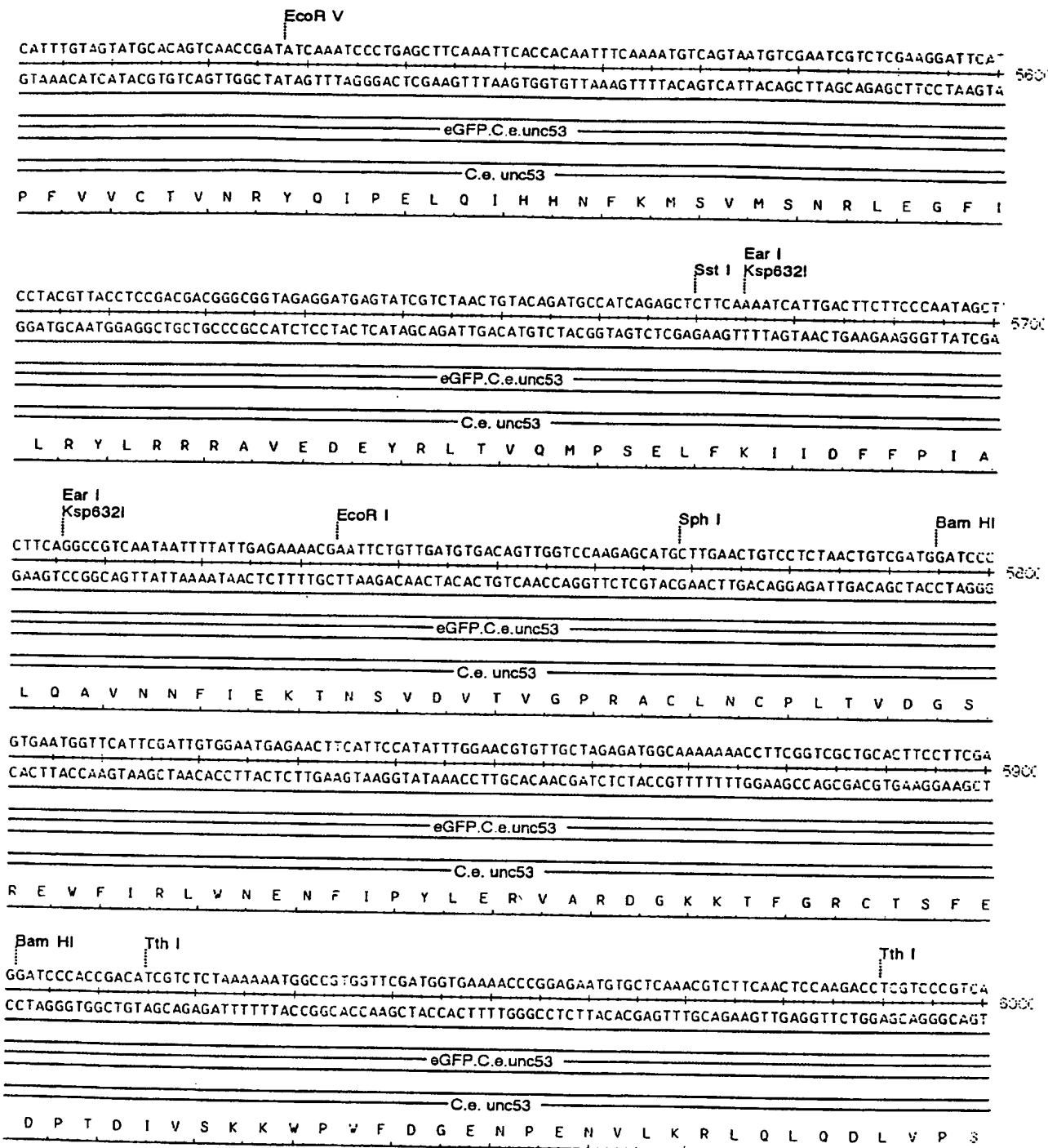
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 10



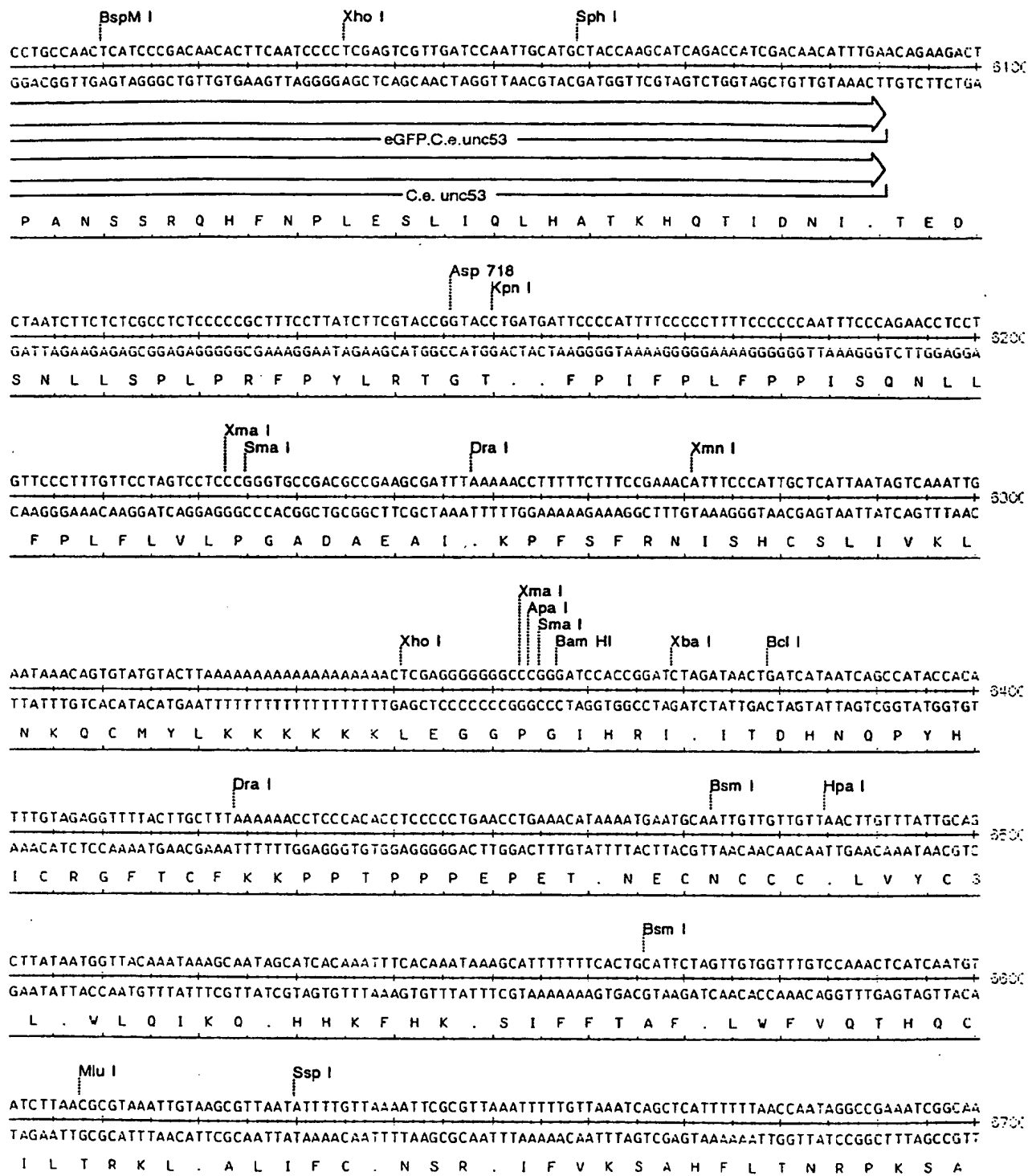
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 11



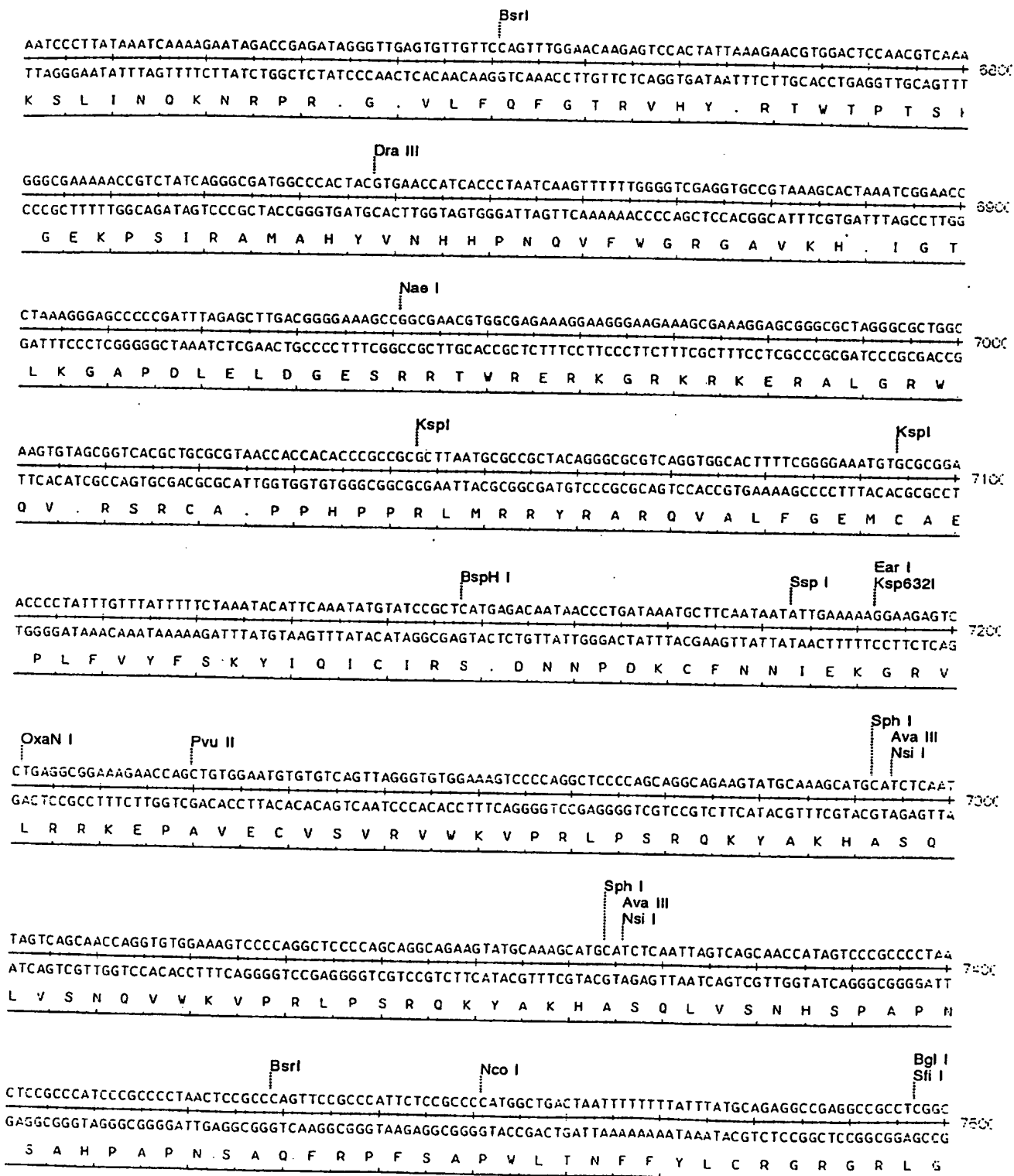
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 12



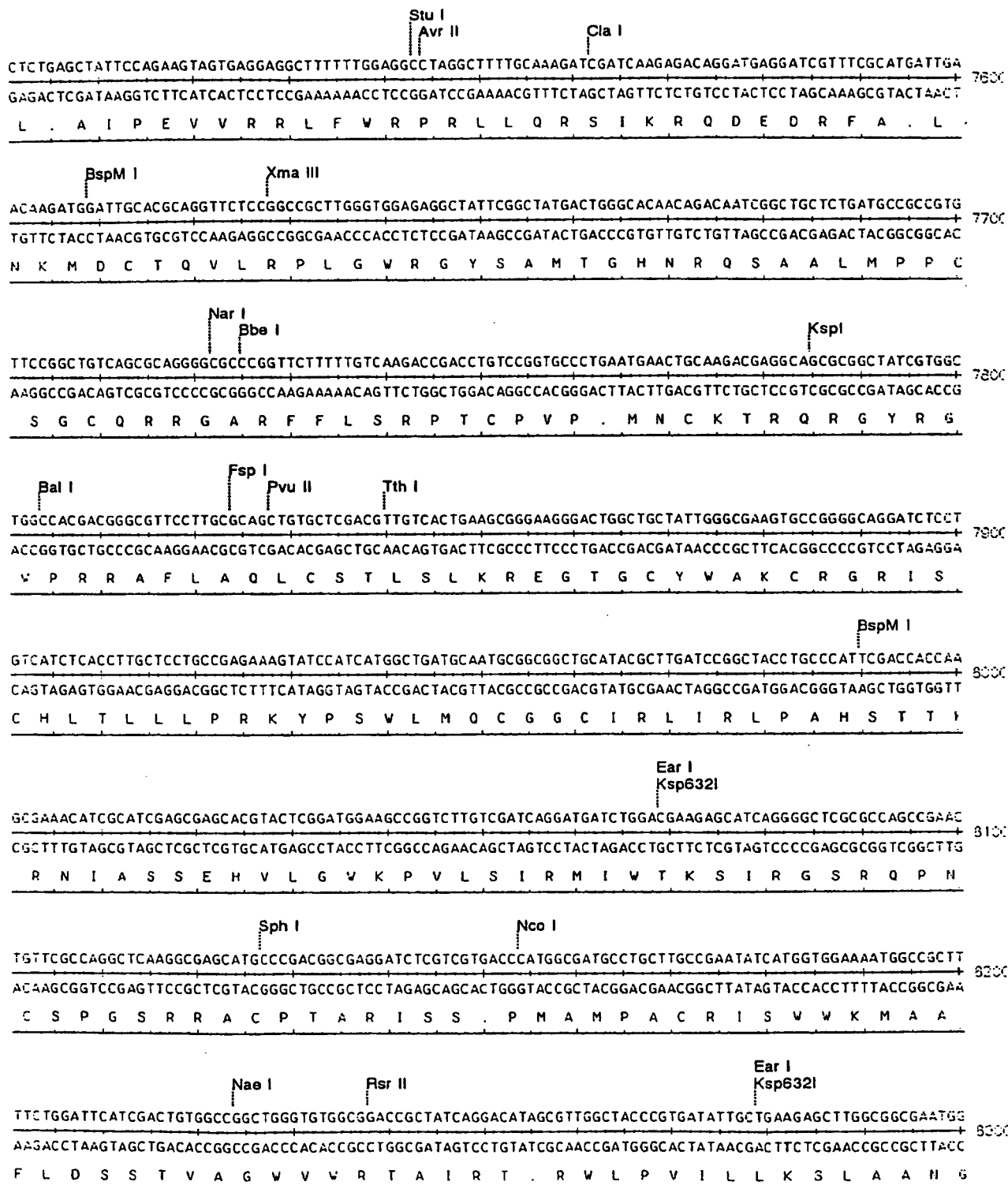
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 13



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 14



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 14

GCTGACCGCTTCTCGTGTCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTCTTCTGAGCGGGACTCTGGG
CGACTGGCGAAGGAGCAGGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAAC TGCTCAAGAAGACTCGCCTGAGACCC 8400
L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G

Asu II BspM I
GTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGG
CAAGCTTTACTGGCTGGTTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCAACCCGAAGCCTTAGCAAAAGGCC 8500
V R N D R P S D A O P A I T R F R F H R R L L . K V G L R N R F P

Nae I KspI Avr II
GACCGCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTTCGCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGG
CTGCGGCGGACCTACTAGGAGGTGCGGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCTCTGTTATGGCC 8600
G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G

KspI
AAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTGGGTCGTTTGTTCATAAACCGGGGTTCCGGTCCCAGGGCTGGCACTC
TTCCTTGGGCGGATATGCGGTTATTTTTCTGTCTTATTTTTCGCTGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAG 8700
R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L

TGTCGATACCCACCGAGACCCATTGGGGCCAATACGCCGCGTTTCTTCTTTTCCCCACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCA
ACAGCTATGGGGTGGCTCTGGGGTAACCCCGTTATGCGGGCGCAAGAAAGGAAAAGGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGT 8800
C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A R

AlwI OxaNI Dra I Dra I
GCCAACGTCGGGGCGGCGAGCCCTGCCATAGCCTCAGGTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTTAAAGGATCTAGGTGA
CGGTTGCAGCCCGCGCTCCGGGACGGTATCGGAGTCAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACT 8900
S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E

BspH I
AGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCTAGGAAAAATATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCTAGAAGAACTCT 9000
D P F . . S H D Q N P L T . V F V P L S V R P R R K D Q R I F L R

TCCTTTTTTCTGCGGTAATCTGCTGCTTGCAAAACAAAAAACCCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCG
AGGAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTGGTGGCGATGGTCCGACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGC 9100
S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S

BsrI
AAGGTAACGGCTTCAGCAGAGCGCAGATACCAATACTGTCTTCTAGTGATAGCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACAT
TTCCATTGACCGAAGTCGTCGCGTCTATGGTTTATGACAGGAAGATCAGATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTA 9200
E G N V L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y I



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 5/10, 15/85, C07K 14/435, 16/18, A61K 38/17, 49/00, C12Q 1/02, G01N 33/53	A3	(11) International Publication Number: WO 98/24810 (43) International Publication Date: 11 June 1998 (11.06.98)
(21) International Application Number: PCT/EP97/06956 (22) International Filing Date: 3 December 1997 (03.12.97) (30) Priority Data: 9625283.8 4 December 1996 (04.12.96) GB (71) Applicant (for all designated States except US): JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B-9032 Wondelgem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Travessera de les Corts, 171/702a, E-08028 Barcelona (ES). DERAEMYAEKER, Marc [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VERHASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F-34000 Montpellier (FR). MAERTENS, Luc, Jacques, Simon [BE/BE]; Vier Uitersten 26, B-8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30,		B-2340 Beerse (BE). VANDEKERCKHOVE, Joel, Stefaan [BE/BE]; Rode Boukendreef 27, B-8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE). (74) Agent: BALDOCK, Sharon, Claire; Boulton Wade Tennant, 27 Fumival Street, London EC4A 1PQ (GB). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 21 January 1999 (21.01.99)
(54) Title: VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS		
(57) Abstract Vertebrate protein homologues of UNC-53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 97/06956

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C12N5/10 C12N15/85 C07K14/435 C07K16/18
A61K38/17 A61K49/00 C12Q1/02 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	WO 96 38555 A (BOGAERT THIERRY (BE); STRINGHAM EVE (CA); VANDEKERCKHOVE JOEL (BE)) 5 December 1996	1, 2, 24-26, 28, 30-36, 38, 40, 42-51, 66, 68, 70, 72, 78, 79, 83-87, 98-105, 107, 112
P, Y	see page 2, line 18 - page 20, line 26	3-23, 27, 29, 37, 39, 41, 58, 59, 67, 69, 74, 80-82,
	-/--	

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

2 October 1998

Date of mailing of the international search report

27. 11. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Donath, C

INTERNATIONAL SEARCH REPORT

Application No
PCT/EP 97/06956

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>see figure 5 see Sequence Listing: SEQ ID NO.: 1 and 4</p> <p style="text-align: center;">---</p> <p>HEKIMI, S. AND KERSHAW, D.: "Axonal guidance defects in a Caenorhabditis elegans mutant reveal cell-extrinsic determinants of neuronal morphology" THE JOURNAL OF NEUROSCIENCE, vol. 13, no. 10, October 1993, pages 4254-4271, XP000612286</p> <p>see page 4254, left-hand column, paragraph 1 - page 4255, right-hand column, paragraph 2 see page 4267, right-hand column, paragraph 1 - page 4271, left-hand column, paragraph 2</p>	<p>88, 115-118, 124,125</p> <p>3-23,27, 29,37, 39,41, 58,59, 67,69, 74, 80-82, 88, 115-118, 124,125</p>
	<p style="text-align: center;">---</p> <p>A STERN, M.J. ET AL.: "The human GRB2 and Drosophila Drk genes can functionally replace the Caenorhabditis elegans cell signaling gene sem-5" MOLECULAR BIOLOGY OF THE CELL, vol. 4, no. 11, November 1993, pages 1175-1188, XP002079466 cited in the application see the whole document</p> <p style="text-align: center;">-----</p>	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 97/06956

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 49-51, 58, 59, 107, 115 and 118 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 52-57, 60-65, 71, 73, 75-77, 89-94, 108-111, 113, 114
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 52-57,60-65,71,73,75-77,89-94,108-111,113,114

Claims 52-57, 60-65, 75-77, 91-94, 108-111,113 and 114 concern a compound, pharmaceutical composition, a nucleic acid sequence or the use of a compound. These compounds or compositions, however, are only defined by the method which can be used in order to identify these compounds or compositions as enhancers or inhibitors of regulation of cell shape, cell growth or motility or of the direction of cell migration or of the signal transduction pathway. Since it is completely unclear which kind of substances will be identified by the respective method and since in the specification no concrete examples for these kind of substances are given, the scope of said claims is totally ambiguous and undefined. Moreover, it cannot be excluded that even substances known in the art may be recognized as an enhancing or inhibiting compound by the respective used method.

The same applies to claims 71, 73, 89 and 90 concerning methods comprising proteins which are only defined by a reference to another method which was used to identify said proteins.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 97/06956

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9638555 A	05-12-1996	AU 6123496 A EP 0832222 A	18-12-1996 01-04-1998
<hr/>			

